Generic properties of subgroups of free groups and finite presentations

Frédérique Bassino, Cyril Nicaud, Pascal Weil

Generic properties of subgroups of free groups and finite presentations

Frédérique Bassino, Cyril Nicaud, and Pascal Weil

Abstract. Asymptotic properties of finitely generated subgroups of free groups, and of finite group presentations, can be considered in several fashions, depending on the way these objects are represented and on the distribution assumed on these representations: here we assume that they are represented by tuples of reduced words (generators of a subgroup) or of cyclically reduced words (relators). Classical models consider fixed size tuples of words (e.g. the few-generator model) or exponential size tuples (e.g. Gromov's density model), and they usually consider that equal length words are equally likely. We generalize both the few-generator and the density models with probabilistic schemes that also allow variability in the size of tuples and non-uniform distributions on words of a given length.

Our first results rely on a relatively mild prefix-heaviness hypothesis on the distributions, which states essentially that the probability of a word decreases exponentially fast as its length grows. Under this hypothesis, we generalize several classical results: exponentially generically a randomly chosen tuple is a basis of the subgroup it generates, this subgroup is malnormal and the tuple satisfies a small cancellation property, even for exponential size tuples. In the special case of the uniform distribution on words of a given length, we give a phase transition theorem for the central tree property, a combinatorial property closely linked to the fact that a tuple freely generates a subgroup. We then further refine our results when the distribution is specified by a Markovian scheme, and in particular we give a phase transition theorem which generalizes the classical results on the densities up to which a tuple of cyclically reduced words chosen uniformly at random exponentially generically satisfies a small cancellation property, and beyond which it presents a trivial group.

This paper is part of the growing body of literature on asymptotic properties of subgroups of free groups and of finite group presentations, which goes back at least to the work of Gromov [**10**] and Arzhantseva and Ol'shanskii [**1**]. As in much of the recent literature, the accent is on so-called generic properties, that is, properties whose probability tends to 1 when the size of instances grows to infinity. A theory

1991 *Mathematics Subject Classification.* Primary 20E05, 60J10 ; Secondary 20E07, 05A16, 68Q17.

Key words and phrases. Asymptotic properties, generic properties, random subgroups, random presentations, Markovian automata, malnormality, small cancellation.

The authors acknowledge partial support from ANR projects ANR 2010 Blan 0202_01 Frec, ANR 2012 JCJC JS02-012-0 MealyM and ANR 2010 Blan 0204_07 Magnum, as well as from ERC grant PCG-336983 and the Programme IdEx Bordeaux - CPU (ANR-10-IDEX-03-02).

of genericity and its applications to complexity theory was initiated by Kapovich, Myasnikov, Schupp and Shpilrain [**14**], and developed in a number of papers, see Kapovich for a recent discussion [**13**].

Genericity, and more generally asymptotic properties, depends on the fashion in which input is represented: finitely presented groups are usually given by finite presentations, i.e. tuples of cyclically reduced words; finitely generated subgroups of free groups can be represented by tuples of words (generators) or Stallings graphs. The representation by Stallings graphs is investigated by the authors, along with Martino and Ventura in [**4, 3, 5**] but we will not discuss it in this paper: we are dealing, like most of the literature, with tuples of words.

There are, classically, two main models (see Section 2.2): the few words model, where an integer k is fixed and one considers k-tuples of words of length at most n, when n tends to infinity, see e.g. [**1, 12, 3, 5**]; and the density model, where we consider tuples of cyclically reduced words of length n, whose size grows exponentially with n, see e.g. [**10, 25, 7, 23**].

Typical properties investigated include the following (see in particular Sections 1.2 and 1.3): whether a random tuple $\vec{h}$ freely generates the subgroup $H = \langle \vec{h} \rangle$ [**1, 12**], whether H is malnormal [**12, 3**] or Whitehead minimal [**27, 5**], whether the finite presentation with relators $\vec{h}$ has a small cancellation property, or whether the group it presents is infinite or trivial [**23**].

All these models implicitly assume the uniform distribution on the set of reduced words of equal length (Ollivier also considers non-uniform distributions in [**23**]).

We introduce (Section 3) a model for probability distributions on tuples of reduced words that is sufficiently general to extend the few words model and Gromov's density model mentioned above, and to leave space for non uniform distributions. Like these two models, ours assumes that a tuple $\vec{h}$ of words is generated by independently drawing words of given lengths, but it also handles independently the size of $\vec{h}$ and the lengths of the words in $\vec{h}$.

Our first set of results assumes a *prefix-heaviness* hypothesis on the probability distribution on words: the probability of drawing a word decreases exponentially fast as its length grows (precise definitions are given in Section 3). It is a natural hypothesis if we imagine that our probabilistic source generates words one letter at a time, from left to right. This relatively mild hypothesis suffices to obtain general results on the exponential genericity of a certain geometric property of the Stallings graph of the subgroup H generated by a randomly chosen tuple $\vec{h}$ (the *central tree property*, implicitly considered in [**1, 12**] and explicitly in [**5**]), of the fact that $\vec{h}$ freely generates H, and of the malnormality of H, see Section 3.5.

In Section 3.6, we apply these general results to the uniform distribution and generalize known results in two directions. Firstly we consider random exponential size tuples, for which we give a phase transition theorem for the central tree property: it holds exponentially generically up to density $\frac{1}{4}$, and fails exponentially generically at densities greater than $\frac{1}{4}$ (Proposition 3.21). In particular, a random tuple is exponentially generically a basis of the subgroup it generates up to density $\frac{1}{4}$, but we cannot say anything of that property at higher densities.

We also extend Jitsukawa's result on malnormality [**12**], from fixed size to exponential size tuples under uniform distribution up to density $\frac{1}{16}$ (Proposition 3.22).

In view of the methods used to establish this result, it is likely that the value $\frac{1}{16}$ is not optimal.

Secondly, we show that the height of the central tree of a random fixed size tupe $\vec{h}$, which measures the amount of initial cancellation between the elements of $\vec{h}$ and $\vec{h}^{-1}$, is generically less than any prescribed unbounded non-decreasing function (Proposition 3.24). Earlier results only showed that this height was exponentially generically bounded by any linear function.

We then introduce *Markovian automata*, a probabilistic automata-theoretic model, to define explicit instances of prefix-heavy distributions (Section 4). Additional assumptions like irreducibility or ergodicity lead to the computation of precise bounds for the parameters of prefix-heaviness. In particular, we prove a phase transition theorem for ergodic Markovian automata (Section 4.4), showing that small cancellation properties generically hold up to a certain density, and generically do not hold at higher densities. More precisely, if $\alpha_{[2]}$ is the coincidence probability of the Markovian automaton, Property $C'(\lambda)$ holds exponentially generically at $\alpha_{[2]}$-density less than $\frac{\lambda}{2}$ (that is: for random tuples of size $\alpha_{[2]}^{-dn}$ for some $d < \frac{\lambda}{2}$), and fails exponentially generically at $\alpha_{[2]}$-densities greater than $\frac{\lambda}{2}$. We also show that at $\alpha_{[2]}$-densities greater than $\frac{1}{2}$, a random tuple of cyclically reduced words generically presents a degenerate group (see Proposition 4.23 for a precise definition). These results generalize the classical results on uniform distribution in Ollivier [**23, 24**]. It remains to be seen whether our methods can be applied to fill the gap, say, between $\alpha_{[2]}$-density $\frac{1}{12}$ and $\frac{1}{2}$, where small cancellation property $C'(\frac{1}{6})$ generically does not hold yet the presented group might be hyperbolic, see [**23, 24**].

Some of the definitions in this paper, notably that of Markovian automata, were introduced by the authors in [**2**], and some of the results were announced there as well. The results in the present paper are more precise, and subsume those of [**2**].

1. Free groups, subgroups and presentations

In this section, we set the notation and basic definitions of the properties of subgroups of free groups and finite presentations which we will consider.

1.1. Free groups and reduced words. Let A be a finite non-empty set, which will remain fixed throughout the paper, with $|A| = r$, and let $\tilde{A}$ be the symmetrized alphabet, namely the disjoint union of A and a set of formal inverses $A^{-1} = \{a^{-1} \in A \mid a \in A\}$. By convention, the formal inverse operation is extended to $\tilde{A}$ by letting $(a^{-1})^{-1} = a$ for each $a \in A$. A word in $\tilde{A}^*$ (that is: a word written on the alphabet $\tilde{A}$) is *reduced* if it does not contain length 2 factors of the form aa^{-1} ($a \in \tilde{A}$). If a word is not reduced, one can *reduce* it by iteratively deleting every factor of the form aa^{-1}. The resulting reduced word is uniquely determined: it does not depend on the order of the cancellations. For instance, $u = aabb^{-1}a^{-1}$ reduces to aaa^{-1}, and thence to a.

The set F of reduced words is naturally equipped with a group structure, where the product $u \cdot v$ is the (reduced) word obtained by reducing the concatenation uv. This group is called the *free group* on A. More generally, every group isomorphic to F, say, $G = \varphi(F)$ where φ is an isomorphism, is said to be a free group, *freely generated by* $\varphi(A)$. The set $\varphi(A)$ is called a *basis* of G. Note that if $r \geq 2$, then F has infinitely many bases: if, for instance, $a \neq b$ are elements of A, then replacing

a by $b^n a b^m$ (for some integers n, m) yields a basis. The *rank* of F (or of any isomorphic free group) is the cardinality $|A|$ of A, and one shows that this notion is well-defined in the following sense: every basis of F has the same cardinality.

Let x, y be elements of a group G. We say that y is a *conjugate* of x if there exists an element $g \in G$ such that $y = g^{-1}xg$, which we write $y = x^g$. The notation is extended to subsets of G: if $H \subseteq G$, then $H^g = \{x^g \mid x \in H\}$. Conjugacy of elements of the free group F is characterized as follows. Say that a word u is *cyclically reduced word* if it is non-empty, reduced and its first and last letters are not mutually inverse (or equivalently, if u^2 is non-empty and reduced). For instance, $ab^{-1}a^{-1}bbb$ is cyclically reduced, but $ab^{-1}a^{-1}bba^{-1}$ is not.

For every reduced word u, let $\kappa(u)$ denote its *cyclic reduction*, which is the shortest word v such that $u = wvw^{-1}$ for some word w. For instance, $\kappa(ab^{-1}a^{-1}bba^{-1}) = a^{-1}b$. It is easily verified that two reduced words u and v are conjugates if and only if $\kappa(u)$ and $\kappa(v)$ are *cyclic conjugates* (that is: there exist words x and y such that $\kappa(u) = xy$ and $\kappa(v) = yx$).

Let $\mathcal{R}_n$ (resp. $\mathcal{C}_n$) denote the set of all reduced (resp. cyclically reduced) words of length $n \geq 1$, and let $\mathcal{R} = \bigcup_{n\geq 1} \mathcal{R}_n$ and $\mathcal{C} = \bigcup_{n\geq 1} \mathcal{C}_n$ be the set of all reduced words, and all cyclically reduced words, respectively.

Every word of length 1 is cyclically reduced, so $|\mathcal{R}_1| = |\mathcal{C}_1| = 2r$. A reduced word of length $n \geq 2$ is of the form ua, where u is reduced and a is not the inverse of the last letter of u. An easy induction shows that there are $|\mathcal{R}_n| = 2r(2r-1)^{n-1} = \frac{2r}{2r-1}(2r-1)^n$ reduced words of length $n \geq 2$.

Similarly, if $n \geq 2$, then $\mathcal{C}_n$ is the set of words of the form ua, where u is a reduced word and $a \in \tilde{A}$ is neither the inverse of the first letter of u, nor the inverse of its last letter: for a given u, there are either $2r-1$ or $2r-2$ such words, depending whether the first and last letter of u are equal. In particular, the number of words in $\mathcal{C}_n$ satisfies $\frac{2r}{2r-1}(2r-1)^{n-1}(2r-2) \leq |\mathcal{C}_n| \leq \frac{2r}{2r-1}(2r-1)^n$, and in particular, $|\mathcal{C}_n| = \Theta((2r-1)^n)$.

1.2. Subgroups and presentations. Given a tuple $\vec{h} = (h_1, \ldots, h_k)$ of elements of F, let $\vec{h}^{\pm} = (h_1, h_1^{-1}, \ldots, h_k, h_k^{-1})$ and let $\langle\vec{h}\rangle$ denote the subgroup of F generated by the elements of $\vec{h}$, that is, the set of all the elements of F which can be written as a product of elements of $\vec{h}^{\pm}$. It is a classical result of Nielsen that every such subgroup is free [**22**].

An important property of subgroups is malnormality, which is related to geometric considerations (*e.g.* [**9, 17**]): a subgroup H of a group G is *malnormal* if $H \cap H^x$ is trivial for every $x \notin H$. It is decidable whether a finitely generated subgroup $\langle\vec{h}\rangle$ is malnormal ([**12, 15**], see Section 1.3), whereas malnormality is not decidable in general hyperbolic groups [**6**].

A tuple $\vec{h}$ of elements of $F(A)$ can also be considered as a set of relators in a group presentation. More precisely, we denote by $\langle A \mid \vec{h}\rangle$ the group with generator set A and relators the elements of $\vec{h}$, namely the quotient of $F(A)$ by the normal subgroup generated by $\vec{h}$. It is customary to consider such a group presentation only when $\vec{h}$ consists only of cyclically reduced words, since $\langle A \mid \vec{h}\rangle = \langle A \mid \kappa(\vec{h})\rangle$.

The small cancellation property is a combinatorial property of a group presentation, with far-reaching consequences on the quotient group. Let $\vec{h}$ be a tuple of cyclically reduced words. A *piece* in $\vec{h}$ is a word u with at least two occurrences as a prefix of a cyclic conjugate of a word in $\vec{h}^{\pm}$. Let $0 < \lambda < 1$. The tuple $\vec{h}$ (or the

group presentation $\langle A \mid \vec{h}\rangle$) has the *small cancellation property* $C'(\lambda)$ if whenever a piece u occurs as a prefix of a cyclic conjugate w of a word in $\vec{h}^{\pm}$, then $|u| < \lambda|w|$.

The following properties are well-known. We do not give the definition of the group-theoretic properties in this statement and refer the reader to [**19**] or to the comprehensive survey [**24**].

PROPOSITION 1.1. *If $\vec{h}$ is a tuple of cyclically reduced words satisfying $C'(\frac{1}{6})$, then $G = \langle A \mid \vec{h}\rangle$ is infinite, torsion-free and word-hyperbolic. In addition, it has solvable word problem (by Dehn's algorithm) and solvable conjugacy problem.*

Moreover, if $\vec{h}$ and $\vec{g}$ both have property $C'(\frac{1}{6})$ and if they present the same group, then $\vec{h}^{\pm} = \vec{g}^{\pm}$ up to the order of the elements in the tuples.

1.3. Graphical representation of subgroups and the central tree property. A privileged tool for the study of subgroups of free groups is provided by *Stallings graphs*: if H is a finitely generated subgroup of F, its Stallings graph $\Gamma(H)$ is a finite graph of a particular type, uniquely representing H, whose computation was first made explicit by Stallings [**31**]. The mathematical object itself is already described by Serre [**29**]. The description we give below differs slightly from Serre's and Stallings', it follows [**35, 15, 33, 21, 30**] and it emphasizes the combinatorial, graph-theoretical aspect, which is more conducive to the discussion of algorithmic properties.

A *finite A-graph* is a pair $\Gamma = (V, E)$ with V finite and $E \subseteq V \times A \times V$, such that if both (u, a, v) and (u, a, v') are in E then $v = v'$, and if both (u, a, v) and (u', a, v) are in E then $u = u'$. Let $v \in V$. The pair (Γ, v) is said to be *admissible* if the underlying graph of Γ is connected (that is: the undirected graph obtained from Γ by forgetting the letter labels and the orientation of edges), and if every vertex $w \in V$, except possibly v, occurs in at least two edges in E.

Every admissible pair $(\Gamma, 1)$ represents a unique subgroup H of $F(A)$ in the following sense: if u is a reduced word, then $u \in H$ if and only if u labels a loop at 1 in Γ (by convention, an edge (u, a, v) can be read from u to v with label a, or from v to u with label a^{-1}). One can show that H is finitely generated. More precisely, the following procedure yields a basis of H: choose a spanning tree T of Γ; for each edge $e = (u, a, v)$ of Γ not in T, let $b_e = x_u a x_v^{-1}$, where x_u (resp. x_v) is the only reduced word labeling a path in T from 1 to u (resp. v); then the b_e freely generate H and as a result, the rank of H is exactly $|E| - |V| + 1$.

Conversely, if $\vec{h} = (h_1, \dots, h_k)$ is a tuple of reduced words, then the subgroup $H = \langle \vec{h}\rangle$ admits a Stallings graph, written $(\Gamma(H), 1)$, which can be computed effectively and efficiently. A quick description of the algorithm is as follows. We first build a graph with edges labeled by letters in $\tilde{A}$, and then reduce it to an A-graph using *foldings*. First build a vertex 1. Then, for every $1 \leq i \leq k$, build a loop with label h_i from 1 to 1, adding $|h_i| - 1$ new vertices. Change every edge (u, a^{-1}, v) labeled by a letter of A^{-1} into an edge (v, a, u). At this point, we have constructed the so-called *bouquet of loops* labeled by the h_i.

Then iteratively identify the vertices v and w whenever there exists a vertex u and a letter $a \in A$ such that either both (u, a, v) and (u, a, w) or both (v, a, u) and (w, a, u) are edges in the graph (the corresponding two edges are *folded*, in Stallings' terminology).

The resulting graph Γ is such that $(\Gamma, 1)$ is admissible, the reduced words labeling a loop at 1 are exactly the elements of H and, very much like in the (1-dimensional) reduction of words, that graph does not depend on the order used to perform the foldings. The graph $(\Gamma(H), 1)$ can be computed in time almost linear (precisely: in time $\mathcal{O}(n \log^* n)$ [**33**]).

Some algebraic properties of H can be directly seen on its Stallings graph $(\Gamma(H), 1)$. For instance, one can show that H is malnormal if and only if there exists no non-empty reduced word u which labels a loop in two distinct vertices of $\Gamma(H)$ [**12, 15**]. This property leads to an easy decision procedure of malnormality for subgroups of a free group. We refer the reader to [**31, 35, 15, 21**] for more information about Stallings graphs.

If $\vec{h}$ is a tuple of elements of F, let $\mathsf{min}(\vec{h})$ be the minimum length of an element of $\vec{h}$ and let $\mathsf{lcp}(\vec{h})$ be the length of the longest common prefix between two words in $\vec{h}^{\pm}$, see Figure 1[1]. We say that $\vec{h}$ has the *central tree property* if $2\,\mathsf{lcp}(\vec{h}) < \mathsf{min}(\vec{h})$.

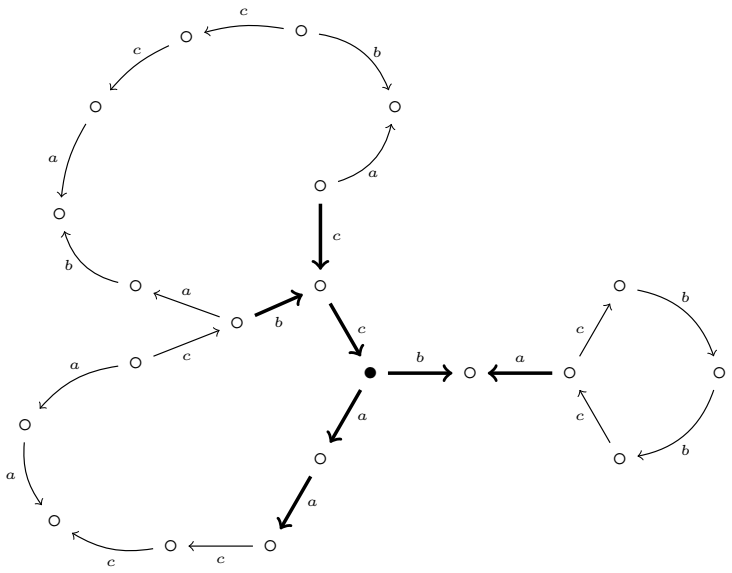

FIGURE 1. The Stallings graph of the subgroup generated by $\vec{h} = (ba^{-1}cb^2a^2b^{-1}, a^2c^2a^{-2}cbc, c^{-1}b^{-1}aba^{-1}c^{-2}ba^{-1}c^2)$, has the central tree property and satisfies $\mathsf{lcp}(\vec{h}) = 2$. The origin is denoted by $\bullet$ and the central tree is depicted in bold arrows.

PROPOSITION 1.2. *Let $\vec{h} = (h_1, \ldots, h_k)$ be a tuple of elements of $F(A)$ with the central tree property and let $H = \langle \vec{h} \rangle$. Then the Stallings graph $\Gamma(H)$ consists of a* central tree *of height $t = \mathsf{lcp}(\vec{h})$ and of k* outer loops*, one for each h_i, connecting*

[1]This definition is closely related with the notion of *trie* of $\vec{h}^{\pm}$. The height of the trie of $\vec{h}^{\pm}$ is $1 + \mathsf{lcp}(\vec{h})$.

the length t prefix and the length t suffix of h_i (two leaves of the central tree), of length $|h_i| - 2t$ respectively. The set of vertices of the central tree can be identified with the set of prefixes of length at most t of the words of $\vec{h}^{\pm}$.

In particular, $\vec{h}$ is a basis of H. Moreover, if $\vec{g}$ is a basis of H also with the central tree property, then $\vec{h}^{\pm}$ and $\vec{g}^{\pm}$ coincide up to the order of their elements.

PROOF. The central tree property shows that the cancellation (folding) that occurs when one considers the bouquet of h_i-labeled loops around the origin, stops before canceling entirely any one of the h_i. The result follows immediately. □

Under the central tree property, we record an interesting sufficient condition for malnormality.

PROPOSITION 1.3. *Let $\vec{h} = (h_1, \dots, h_k)$ be a tuple of elements of $F(A)$ with the central tree property and let $H = \langle \vec{h} \rangle$. Let us assume additionally that $3\,\mathsf{lcp}(\vec{h}) < \mathsf{min}(\vec{h})$ and that no word of length at least $\frac{1}{2}(\mathsf{min}(\vec{h}) - 3\,\mathsf{lcp}(\vec{h}))$ has several occurrences as a factor of an element of $\vec{h}^{\pm}$, then H is malnormal.*

REMARK 1.4. In the proof below, and in several other statements and proofs later in the paper, we consider words whose length is specified by an algebraic expression which does not always compute to an integer (*e.g.*, $\frac{1}{2}(\mathsf{min}(\vec{h}) - 3\,\mathsf{lcp}(\vec{h}))$). To be rigorous, we should consider only the integer part of these expressions. For the sake of simplicity, we dispense with this extra notation, and implicitly consider that if a word of length ℓ is considered, then we mean that its length is $\lfloor \ell \rfloor$.

PROOF. Let $m = \mathsf{min}(\vec{h})$ and $t = \mathsf{lcp}(\vec{h})$. Proposition 1.2 shows that $\Gamma(H)$ consists of a central tree of height t and of outer loops, one for each h_i, of length $|h_i| - 2t \geq m - 2t$.

If H is not malnormal, then a word u labels a loop at two distinct vertices of $\Gamma(H)$. Without loss of generality, u is cyclically reduced. Moreover, given the particular geometry of $\Gamma(H)$, both loops visit the central tree. Without loss of generality, we may assume that one of the u-labeled loops starts in the central tree, at distance exactly t from the base vertex 1, and travels away from 1. In particular, $|u| \geq m - 2t$, and if v is the prefix of u of length $m - 2t$, then v is a factor of some $h_i^{\pm 1}$.

Let s be the start state of the second u-labeled loop: reading this loop starts with reading the word v. Suppose that s is in the central tree: either reading u (and v) from s takes us away from 1 towards a leaf of the central tree and into an outer loop, and v is a factor of some $h_j^{\pm 1}$; or reading v from s moves us towards 1 for a distance at most t, after which the path travels away from 1, along a path labeled by a factor of some $h_j^{\pm 1}$, for a distance at least $m - 3t$. In either case, a factor of u of length $m - 3t > \frac{1}{2}(m - 3t)$ has two occurrences in $\vec{h}^{\pm}$.

Suppose now that s is on an outer loop (say, associated to $h_j^{\pm 1}$) and that s' is the first vertex of the central tree reached along the loop. If s' is reached after reading a prefix of u of length greater than $\frac{1}{2}(m - 3t)$, then the prefix of v of length $\frac{1}{2}(m - 3t)$ is a factor of $h_j^{\pm 1}$. Otherwise v labels a path from s which first reaches s', then travels towards 1 in the central tree for a distance at most t, and thence away from 1, along a path labeled by some $h_\ell^{\pm 1}$, which it follows over a length at least equal to $(m - 2t) - \frac{1}{2}(m - 3t) - t = \frac{1}{2}(m - 3t)$.

Thus, in every case, u contains a factor of length $\frac{1}{2}(m-3t)$ with two distinct occurrences as a factor of an element of $\vec{h}^{\pm}$ and this concludes the proof. □

To conclude this section, we note that the properties discussed above are preserved when going from a tuple $\vec{h}$ to a sub-tuple: say that a tuple $\vec{g}$ is *contained in* a tuple $\vec{h}$, written $\vec{g} \leq \vec{h}$, if every element of $\vec{g}$ is an element of $\vec{h}$.

PROPOSITION 1.5. *Let $\vec{g}, \vec{h}$ be tuples of reduced words such that $\vec{g} \leq \vec{h}$.*

- *If $\vec{h}$ has the central tree property, so does $\vec{g}$.*
- *If $\vec{h}$ consists of cyclically reduced words and $\vec{h}$ has Property $C'(\lambda)$, then so does $\vec{g}$.*
- *If $\vec{h}$ has the central tree property, then $\langle \vec{g} \rangle$ is a free factor of $\langle \vec{h} \rangle$, and $\langle \vec{g} \rangle$ is malnormal if $\langle \vec{h} \rangle$ is.*

PROOF. The first two properties are immediate from the definition. Supose now that $\vec{h}$ has the central tree property. Then by Proposition 1.2, $\vec{h}$ is a basis of $\langle \vec{h} \rangle$, and by the first statement of the current proposition, $\vec{g}$ is a basis of $\langle \vec{g} \rangle$. Since $\vec{g} \leq \vec{h}$, $\langle \vec{g} \rangle$ is a free factor of $\langle \vec{h} \rangle$.

In particular, $\langle \vec{g} \rangle$ is malnormal in $\langle \vec{h} \rangle$ (a free factor always is, by elementary reasons). It is immediate from the definition that malnormality is transitive, so if $\langle \vec{h} \rangle$ is malnormal in F, then so is $\langle \vec{g} \rangle$. □

2. Random models and generic properties

We will discuss several models of randomness for finitely presented groups and finitely generated subgroups, or rather, for finite tuples of cyclically reduced words (group presentations) and finite tuples of reduced words. In this section, we fix a general framework for these models of randomness and we survey some of the known results.

2.1. Generic properties and negligible properties. Let us say that a function f, defined on $\mathbb{N}$ and such that $\lim f(n) = 0$, is *exponentially* (resp. *super-polynomially*, *polynomially*) *small* if $f(n) = o(e^{-dn})$ for some $d > 0$ (resp. $f(n) = o(n^{-d})$ for every positive integer d, $f(n) = o(n^{-d})$ for some positive integer d).

Given a sequence of probability laws $(\mathbb{P}_n)_n$ on a set S, we say that a subset $X \subseteq S$ is *negligible* if $\lim_n \mathbb{P}_n(X) = 0$, and *generic* if its complement is negligible.[2]

We also say that X is *exponentially* (resp. *super-polynomially*, *polynomially*) *negligible* if $\mathbb{P}_n(X)$ tends to 0 and is exponentially (resp. super-polynomially, polynomially) small. And it is *exponentially* (resp. *super-polynomially*, *polynomially*) *generic* if its complement is exponentially (resp. super-polynomially, polynomially) negligible.

In this paper, the set S will be the set of all finite tuples of reduced words, or cyclically reduced words, and the probability laws $\mathbb{P}_n$ will be such that every subset is measurable: we will therefore not specify in the statements that we consider only measurable sets.

The notions of genericity and negligibility have elementary closure properties that we will use freely in the sequel. For instance, a superset of a generic set is

[2]This is the same notion as *with high probability* or *with overwhelming probability*, which are used in the discrete probability literature.

generic, as well as the intersection of finitely many generic sets. Dual properties hold for negligible sets.

2.2. The few-generator model and the density model. In this section, we review the results known on two random models, originally introduced to discuss finite presentations. We discuss more general models in Section 3 below.

2.2.1. *The few-generator model.* In the *few-generator model*, an integer $k \geq 1$ is fixed, and we let $\mathbb{P}_n$ be the uniform probability on the set of k-tuples of words of F of length at most n. Proposition 2.1 is established by elementary counting arguments, see Gromov [**10**, Prop. 0.2.A] or Arzhantseva and Ol'shanskii [**1**, Lemma 3].

PROPOSITION 2.1. *Let $k \geq 1$, $0 < \alpha < \frac{1}{2}$, $2\alpha < \beta < 1$ and $0 < \lambda < 1$. Then a k-tuple $\vec{h}$ of elements of F of length at most n picked uniformly at random, exponentially generically satisfies the following properties:*

- $\mathsf{min}(\vec{h}) > \beta n$,
- $\mathsf{lcp}(\vec{h}) < \alpha n$,
- *no word of length λn has two occurrences as a factor of an element of $\vec{h}^{\pm}$.*

In view of Propositions 1.2 and 1.3, this yields the following corollary ([**3**], and [**12**] for the malnormality statement).

COROLLARY 2.2. *Let $k \geq 1$. If $\vec{h}$ is a k-tuple of elements of F of length at most n picked uniformly at random and $H = \langle \vec{h} \rangle$, then*

- *exponentially generically, $\vec{h}$ has the central tree property, and in particular, $\Gamma(H)$ can be constructed in linear time (in $k \cdot n$), simply by computing the initial cancellation of the elements of $\vec{h}^{\pm}$; H is freely generated by the elements of $\vec{h}$, and H has rank k;*
- *exponentially generically, H is malnormal.*

Moreover, if $\vec{h}$ and $\vec{g}$ generate the same subgroup, then exponentially generically, $\vec{h}^{\pm} = \vec{g}^{\pm}$ up to the order of the elements in the tuples.

The following statement follows from Proposition 1.5, and from Theorem 2.4 below (which is independent).

COROLLARY 2.3. *In the few-generator model, if $\vec{h}$ is a k-tuple of cyclically reduced words of length at most n, then*

- *for any $0 < \lambda < \frac{1}{2}$, $\vec{h}$ exponentially generically satisfies the small cancellation property $C'(\lambda)$;*
- *exponentially generically, the group $\langle A \mid \vec{h} \rangle$ is infinite, torsion-free, word-hyperbolic, it has solvable word problem (by Dehn's algorithm) and solvable conjugacy problem.*

2.2.2. *The density model.* In the *density model*, a density $0 < d < 1$ is fixed, and a tuple of cyclically reduced elements of the n-sphere of density d is picked uniformly at random: that is, the tuple $\vec{h}$ consists of $|\mathcal{C}_n|^d$ cyclically reduced words of length n. This model was introduced by Gromov [**11**] and complete proofs were given by Ol'shanskii [**25**], Champetier [**7**] and Ollivier [**23**].

THEOREM 2.4. *Let $0 < \alpha < d < \beta < 1$. In the density model, the following properties hold:*

(1) *exponentially generically, every word of length αn occurs as a factor of a word in $\vec{h}$, and some word of length βn fails to occur as a factor of a word in $\vec{h}^{\pm}$;*
(2) *if $d < \frac{1}{2}$, then exponentially generically, $\vec{h}$ satisfies property $C'(\lambda)$ for $\lambda > 2d$ but $\vec{h}$ does not satisfy $C'(\lambda)$ for $\lambda < 2d$; in particular, at density $d < \frac{1}{12}$, $\vec{h}$ satisfies exponentially generically property $C'(\frac{1}{6})$ and the group $\langle A \mid \vec{h}\rangle$ is infinite and hyperbolic; and at density $d > \frac{1}{12}$, exponentially generically, $\vec{h}$ does not satisfy $C'(\frac{1}{6})$;*
(3) *at density $d > \frac{1}{2}$, exponentially generically, $\langle \vec{h}\rangle$ is equal to $F(A)$, or has index 2. In particular, the group $\langle A \mid \vec{h}\rangle$ is either trivial or $\mathbb{Z}/2\mathbb{Z}$;*
(4) *at density $d < \frac{1}{2}$, the group $\langle A \mid \vec{h}\rangle$ is generically infinite and hyperbolic.*

Properties (1)-(3) in Theorem 2.4 are obtained by counting arguments. Property (4) is the "hard part" of the theorem, where hyperbolicity does not follow from a small cancellation property.

As pointed out by Ollivier [**24**, Sec. I.2.c], the statement of Theorem 2.4 still holds if a tuple of cyclically reduced elements is chosen uniformly at random at density d in the n-ball rather than in the n-sphere (that is, it consists of words of length at most n). We will actually verify this fact again in Section 3.6.

3. A general probabilistic model

We introduce a fairly general probabilistic model, which generalizes both the few-generator and the density models.

3.1. Prefix-heavy sequences of measures on reduced words. For every reduced word $u \in \mathcal{R}$, let $\mathcal{P}(u)$ be the set of all reduced words v of which u is a prefix (that is: $\mathcal{P}(u) = u\tilde{A}^* \cap \mathcal{R}$). Let also $\mathcal{P}_n(u)$ be the set $\mathcal{R}_n \cap \mathcal{P}(u)$. The notation $\mathcal{P}$ can also be extended to a set U of reduced words: $\mathcal{P}(U) = \bigcup_{u\in U} \mathcal{P}(u)$.

Let $(\mathbb{R}_n)_{n\geq 0}$ be a sequence of probability measures on $\mathcal{R}$ and let $C \geq 1$ and $\alpha \in (0, 1)$. We say that the sequence $(\mathbb{R}_n)_{n\geq 0}$ is a *prefix-heavy sequence of measures on $\mathcal{R}$* of parameters (C, α) if:

(1) for every $n \geq 0$, the support of the measure $\mathbb{R}_n$ is included in $\mathcal{R}_n$;
(2) for every $n \geq 0$ and for every $u \in \mathcal{R}$, if $\mathbb{R}_n(\mathcal{P}(u)) \neq 0$ then for every $v \in \mathcal{R}$

$$\mathbb{R}_n\big(\mathcal{P}(uv) \mid \mathcal{P}(u)\big) \leq C\alpha^{|v|}.$$

This prefix-oriented definition is rather natural if one thinks of a source as generating reduced words from left to right, as is usual in information theory.

Remark 3.1. Taking $u = \varepsilon$ in the definition yields $\mathbb{R}_n\big(\mathcal{P}(v)\big) \leq C\alpha^{|v|}$. For $n = |v|$, we have $\mathcal{P}(v) \cap \mathcal{R}_n = \{v\}$, so the probability of v decreases exponentially with the length of v.

Example 3.2. The sequence of uniform distributions on $\mathcal{R}_n$ is a prefix-heavy sequence of measures with parameters $C = 1$ and $\alpha = \frac{1}{2r-1}$. Indeed, if u is a reduced word of length at most $n \geq 0$ (for a longer u, $\mathbb{R}_n(\mathcal{P}(u)) = 0$), and if uv is

reduced, we have

$$\mathbb{R}_n\big(\mathcal{P}(uv) \mid \mathcal{P}(u)\big) = \begin{cases} \frac{1}{(2r-1)^{|v|}} & \text{if } |u| + |v| \leq n \text{ and } u \neq \varepsilon, \\ \frac{1}{2r(2r-1)^{|v|-1}} & \text{if } |v| \leq n \text{ and } u = \varepsilon, \\ 0 & \text{otherwise.} \end{cases}$$

Example 3.3. By a similar computation, one verifies that the sequence of uniform distributions on $\mathcal{C}_n$, the cyclically reduced words, is also a prefix-heavy sequence of measures, with parameters $C = \frac{2r-1}{2r-2}$ and $\alpha = \frac{1}{2r-1}$ (see Section 1.1).

For the rest of this section, we fix a sequence of measures $(\mathbb{R}_n)_{n\geq 0}$ on $\mathcal{R}$, which is prefix-heavy with parameters (C, α). All probabilities refer to this sequence, that is: the probability of a subset of $\mathcal{R}_n$ is computed according to $\mathbb{R}_n$.

Remark 3.4. If X and Y are subsets of $\mathcal{R}$, the notation $\mathbb{R}_n(X \mid Y)$ is technically defined only if $\mathbb{R}_n(Y) \neq 0$. To avoid stating cumbersome hypotheses, we adopt the convention that $\mathbb{R}_n(X \mid Y)\, \mathbb{R}_n(Y) = 0$ whenever $\mathbb{R}_n(Y) = 0$.

3.2. Repeated factors in random reduced words. Let us first evaluate the probability of occurrence of prescribed, non-overlapping factors in a reduced word. Let $m \geq 0$, $\vec{v} = (v_1, \ldots, v_m)$ be a vector of non-empty reduced words and $\vec{\imath} = (i_1, \ldots, i_m)$ be a vector of integers. We denote by $E(\vec{v}, \vec{\imath})$ denote the set of reduced words of length n, admitting v_j as a factor at position i_j for every $1 \leq j \leq m$ (if $m = 0$, then $E(\vec{v}, \vec{\imath}) = \mathcal{R}$). If $n \geq 1$, we also write $E_n(\vec{v}, \vec{\imath})$ for $E(\vec{v}, \vec{\imath}) \cap \mathcal{R}_n$.

Lemma 3.5. *Let $\vec{v} = (v_1, \ldots, v_m)$ be a sequence of non-empty reduced words and $\vec{\imath} = (i_1, \ldots, i_m)$ be a sequence of integers satisfying*

$$1 \leq i_1 < i_1 + |v_1| \leq i_2 < i_2 + |v_2| \leq \ldots \leq i_m + |v_m| \leq n.$$

Then the following inequality holds:

$$\mathbb{R}_n\left(E(\vec{v}, \vec{\imath})\right) \leq C^m \alpha^{|v_1 v_2 \cdots v_m|}.$$

In addition, if $m \geq 1$ and $\vec{x} = (v_1, \ldots, v_{m-1})$ and $\vec{\jmath} = (i_1, \ldots, i_{m-1})$, then

$$\mathbb{R}_n(E(\vec{v}, \vec{\imath})) \leq C\alpha^{|v_m|} \mathbb{R}_n(E(\vec{x}, \vec{\jmath})).$$

Proof. The proof is by induction on m and the case $m = 0$ is trivial. We now assume that $m \geq 1$ and that the inequality holds for vectors of length $m-1$. Since $(\mathbb{R}_n)_n$ is prefix-heavy, we have

$$\mathbb{R}_n(\mathcal{P}(uv_m)) = \mathbb{R}_n(\mathcal{P}(uv_m) \mid \mathcal{P}(u))\ \mathbb{R}_n(\mathcal{P}(u)) \leq C\alpha^{|v_m|} \mathbb{R}_n(\mathcal{P}(u))$$

for each u. Since $E(\vec{v}, \vec{\imath}) = \mathcal{P}(E_{i_m - 1}(\vec{x}, \vec{\jmath})v_m)$, summing the previous inequality over all $u \in E_{i_m - 1}(\vec{x}, \vec{\jmath})$ yields

$$\mathbb{R}_n(E(\vec{v}, \vec{\imath})) \leq C\alpha^{|v_m|} \mathbb{R}_n(\mathcal{P}(E_{i_m-1}(\vec{x}, \vec{\jmath}))) = C\alpha^{|v_m|} \mathbb{R}_n(E(\vec{x}, \vec{\jmath}))$$

since $n \geq i_m + |v_m|$. This concludes the proof. □

Corollary 3.6. *Let $v_1, \ldots, v_m$ be non-empty reduced words. The probability that a word of length n admits $v_1, \ldots, v_m$ in that order as non-overlapping factors, is at most $C^m n^m \alpha^{|v_1 \cdots v_m|}$.*

Proof. This is a direct consequence of Lemma 3.5, summing over all possible position vectors. □

We now consider repeated non-overlapping occurrences of factors of a prescribed length.

LEMMA 3.7. *Let $1 \leq i, j, t \leq n$ be such that $i + t \leq j$. The probability that a word of length t occurs (resp. a word of length t and its inverse occur) at positions i and j in a reduced word of length n is at most equal to $C\alpha^t$.*

The probability that a reduced word of length n has two non-overlapping occurrences of a factor of length t (resp. occurrences of a factor of length t and its inverse) is at most equal to $Cn^2\alpha^t$.

PROOF. Let $E_n(t, i, j)$ be the set of reduced words of length n in which the same factor of length t occurs at positions i and j. Then $E_n(t, i, j)$ is the disjoint union of the sets $E_n((v, v), (i, j))$, where v runs over $\mathcal{R}_t$. By Lemma 3.5, we have

$$\mathbb{R}_n(E_n(t, i, j)) = \sum_{v \in \mathcal{R}_t} \mathbb{R}_n(E((v, v), (i, j))) \leq C\alpha^t \sum_{v \in \mathcal{R}_t} \mathbb{R}_n(E((v), (i))) = C\alpha^t,$$

where the last equality is due to the fact that the $E_n((v), (i))$ form a partition of $\mathcal{R}_n$ when v runs over $\mathcal{R}_t$.

The same reasoning applied to the vectors (v, v^{-1}) yields the analogous inequality for words containing non-overlapping occurrences of a word and its inverse.

The last part of the statement follows by summing over all possible values of i and j. □

Applying Lemma 3.7 with $i = 1$ and $j = n - t + 1$, we get the following useful statement.

COROLLARY 3.8. *For every positive integers n, t such that $n > 2t$, the probability that a reduced word $u \in \mathcal{R}_n$ is of the form vwv^{-1}, for some word v of length t, is at most $C\alpha^t$.*

Finally, we also estimate the probability that a word has two overlapping occurrences of a factor. Note that we do not need to consider overlapping occurrences of a word v and its inverse, since a reduced word cannot overlap with its inverse.

LEMMA 3.9. *Let $1 \leq t < n$. The probability that a reduced word of length n has overlapping occurrences of a factor of length t is at most $Cnt\alpha^t$.*

PROOF. If a word v overlaps with itself, more precisely, if $xv = vz$ for some words x, z such that $0 < |x| = |z| < |v|$, then it is a classical result from combinatorics on words that $v = x^s y$ where $s = \left\lfloor \frac{|v|}{|x|} \right\rfloor \geq 1$ and y is the prefix of x of length $|v| - s|x|$ (see Figure 2).

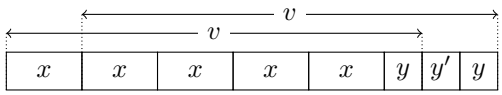

FIGURE 2. A classical result from combinatorics of words: if $xv = vz$ with $0 < |x| < |v|$, then v is of the form $v = x^s y$ for some positive integer s and some prefix y of x.

It follows that, if a reduced word u has (overlapping) occurrences of a factor v of length t at positions i and j ($j < i + t$), then u admits a factor of the form xv at position i, where x is the prefix of v of length $j - i$. Note that, once t and $j - i$ are fixed, v is entirely determined by x. Therefore this occurs with probability

$$P \leq \sum_{i=1}^{n} \sum_{j=i+1}^{i+t-1} \sum_{x \in \mathcal{R}_{j-i}} \mathbb{R}_n(E((xv), (i))) = \sum_{i=1}^{n} \sum_{j=i+1}^{i+t-1} \sum_{x \in \mathcal{R}_{j-i}} \mathbb{R}_n(E((x, v), (i, j))).$$

It follows that

$$P \leq \sum_{i=1}^{n} \sum_{j=i+1}^{i+t-1} C\alpha^t \sum_{x \in \mathcal{R}_{j-i}} \mathbb{R}_n(E((x), (i))) = \sum_{i=1}^{n} \sum_{j=i+1}^{i+t-1} C\alpha^t \leq Cnt\alpha^t.$$

by Lemma 3.5 and using the fact that the $E_n((x), (i))$ form a partition of $\mathcal{R}_n$ when x runs over $\mathcal{R}_{j-i}$. □

3.3. Repeated cyclic factors in random reduced words. A word v is a *cyclic factor* of a word u if either $u \in \tilde{A}^* v \tilde{A}^*$, or $v = v_1 v_2$ and $u \in v_2 \tilde{A}^* v_1$ – in which case we say that v is a *straddling factor*. For now, we only assume that u is reduced, but we will be ultimately interested in the cyclically reduced case, see Corollary 3.14.

LEMMA 3.10. *Let $1 \leq i, t \leq n$ such that $i + t \leq n$ and let v be reduced word v of length t. Then the probability that v is a cyclic factor at position i of an element of $\mathcal{R}_n$, is at most $(Cn + C^2 t)\alpha^t \leq 2C^2 n\alpha^t$.*

PROOF. The probability that v occurs as a (regular) factor of an element of $\mathcal{R}_n$ is at most $Cn\alpha^t$ by Corollary 3.6.

On the other hand, v occurs as a straddling factor of $u \in \mathcal{R}_n$ if $v = v_2 v_1$, with $1 \leq \ell = |v_2| < t$ and $u \in v_1 \tilde{A}^* v_2$, that is, $u \in E((v_1, v_2), (1, n - \ell + 1))$. By Lemma 3.5, this happens with probability at most $C^2\alpha^t$. Summing over the possible values of ℓ, we find that that v occurs as a straddling factor of an element of $\mathcal{R}_n$ with probability at most $C^2 t\alpha^t$.

Therefore the probability that v occurs in u as a cyclic factor is at most $(Cn + C^2 t)\alpha^t$, as announced. □

We now consider multiple occurrences of cyclic factors of a given length.

LEMMA 3.11. *Let $1 \leq t < n$. The probability that a reduced word of length n has two non-overlapping occurrences of a cyclic factor of length t (resp. an occurrence of a cyclic factor of length t and its inverse), is at most $(Cn^2 + C^2 nt)\alpha^t \leq 2C^2 n^2 \alpha^t$.*

PROOF. Again there are several cases, depending whether the occurrences of the word (or the word and its inverse) are both standard factors, or one of them is straddling.

The probability that a reduced word $u \in \mathcal{R}_n$ admits two non-overlapping occurrences of a (standard) factor of length t (resp. occurrences of a factor of length t and its inverse), is at most $Cn^2\alpha^t$ by Lemma 3.7.

We now consider the situation where u has two occurrences of the same word of length t, one as a standard factor and one straddling: there exist integers ℓ, i and reduced words v_1, v_2 such that $0 < \ell < t$, $\ell \leq i \leq n - 2t + \ell$, $|v_2| = \ell$, $|v_1 v_2| = t$ and $u \in E((v_2, v_1 v_2, v_1), (1, i, n - t + \ell + 1)) = E((v_2, v_1, v_2, v_1), (1, i, i + \ell, n - t + \ell + 1))$.

Applying Lemma 3.5 twice, we find that the probability of this event according to $\mathbb{R}_n$ is at most equal to $C^2\alpha^t\mathbb{R}_n(E((v_2,v_1),(1,i)))$.

Then the probability P that a word in $\mathcal{R}_n$ admits two non-overlapping occurrences of a factor of length t, one standard and one straddling, is bounded above by the sum of these values when ℓ, i, v_1, v_2 run over all possible values:

$$P \leq \sum_{\ell=0}^{t} \sum_{i=\ell}^{n-2t+\ell} \sum_{v_2\in\mathcal{R}_\ell} \sum_{v_1\in\mathcal{R}_{t-\ell}} C^2\alpha^t\mathbb{R}_n(E((v_2,v_1),(1,i))).$$

For fixed values of ℓ and i, $\mathcal{R}_n$ is the disjoint union of the $E((v_2,v_1),(1,i))$ when v_2 runs over $\mathcal{R}_\ell$ and v_1 runs over $\mathcal{R}_{t-\ell}$. So we get

$$P \leq \sum_{\ell=0}^{t} \sum_{i=\ell}^{n-2t+\ell} C^2\alpha^t \leq C^2nt\alpha^t.$$

Thus the probability that a reduced word of length n has two non-overlapping occurrences of a word of length t as cyclic factors is at most equal to $(Cn^2 + C^2nt)\alpha^t \leq 2C^2n^2\alpha^t$, as announced.

Finally, we consider the situation where a factor of length t and its inverse occur in u, with one of the occurrences straddling: that is, there exist integers ℓ, i and reduced words v_1, v_2 such that $0 < \ell < t$, $\ell \leq i \leq n-2t+\ell$, $|v_2| = \ell$, $|v_1v_2| = t$ and u lies in

$$E((v_2, v_2^{-1}v_1^{-1}, v_1), (1, i, n-t+\ell+1)) = E((v_2, v_2^{-1}, v_1^{-1}, v_1), (1, i, i+\ell, n-t+\ell+1)).$$

As above, the probability of this event according to $\mathbb{R}_n$ is at most

$$C\alpha^{t-\ell}\mathbb{R}_n(E((v_2, v_2^{-1}, v_1^{-1}), (1, i, i+\ell)))$$

and the probability P' that a reduced word of length n has two non-overlapping occurrences of a word of length t as cyclic factors, with one of them straddling, satisfies

$$P' \leq \sum_{\ell=1}^{t-1} \sum_{i=\ell}^{n-2t+\ell} \sum_{v_2\in\mathcal{R}_\ell} \sum_{v_1\in\mathcal{R}_{t-\ell}} C\alpha^{t-\ell}\mathbb{R}_n(E((v_2, v_2^{-1}, v_1^{-1}), (1, i, i+\ell))).$$

For fixed values of ℓ, i and v_2, $E_n((v_2, v_2^{-1}), (1, i))$ is the disjoint union of the $E_n((v_2, v_2^{-1}, v_1^{-1}), (1, i, i+\ell))$ when v_1 runs over $\mathcal{R}_{t-\ell}$. Therefore we have

$$P' \leq \sum_{\ell=1}^{t-1} \sum_{i=\ell}^{n-2t+\ell} \sum_{v_2\in\mathcal{R}_\ell} C\alpha^{t-\ell}\mathbb{R}_n(E((v_2, v_2^{-1}), (1, i))).$$

By Lemma 3.5 again, $\mathbb{R}_n(E((v_2, v_2^{-1}), (1, i))) \leq C\alpha^\ell\mathbb{R}_n(E((v_2)(1)))$ and we get, by the same reasoning as above,

$$P' \leq \sum_{\ell=1}^{t-1} \sum_{i=\ell}^{n-2t+\ell} \sum_{v_2\in\mathcal{R}_\ell} C^2\alpha^t\mathbb{R}_n(E((v_2),(1))) = \sum_{\ell=1}^{t-1} \sum_{i=\ell}^{n-2t+\ell} C^2\alpha^t \leq C^2nt\alpha^t.$$

Thus the probability that a reduced word of length n has an occurrence of a word of length t and its inverse as a cyclic factor is, again, at most equal to $(Cn^2+C^2nt)\alpha^t \leq 2C^2n^2\alpha^t$, as announced. □

Finally, we give an upper bound to the probability that a reduced word has overlapping occurrences of a cyclic factor of length t (observing again that a reduced word cannot have overlapping occurrences of a (cyclic) factor and its inverse).

LEMMA 3.12. *Let $1 \leq t < n$. The probability that a reduced word of length n has overlapping occurrences of a cyclic factor of length t is at most equal to $\left(Cnt + 2C^2t^2\right)\alpha^t \leq 3\ C^2nt\alpha^t$.*

PROOF. The probability that a reduced word of length n has overlapping occurrences of a non-straddling factor of length t is at most $Cnt\alpha^t$ by Lemma 3.9.

Let us now assume that the reduced word $u \in \mathcal{R}_n$ has overlapping occurrences of a cyclic factor v of length t, with one at least of these occurrences straddling. Note that any cyclic factor of u is a factor of u^2. Therefore, using the same arguments as for Lemma 3.9, u has a straddling cyclic factor of the form $xv = x^{s+1}y$, where $|x| > 0$, y is a prefix of x and $s \geq 1$. In particular, $v = x^s y$ and $t = s|x| + |y|$.

It follows that u is in $v_2\tilde{A}^*v_1$, for some v_1, v_2 such that $v_1v_2 = x^{s+1}y$. Denote by $\mathsf{pref}_\ell(z)$ and $\mathsf{suff}_\ell(z)$ the prefix and the suffix of length ℓ of a word z. Then there exist a cyclic conjugate z of x and integers $0 \leq h, \ell < |z| = |x|$ and $m, m' \geq 0$ such that $v_1 = \mathsf{suff}_h(z)z^{m'}$ and $v_2 = z^m\mathsf{pref}_\ell(z)$. Note that $x^{s+1}y = \mathsf{suff}_h(z)z^{m+m'}\mathsf{pref}_\ell(z)$ and

$$\begin{aligned} h + \ell &= |y| \pmod{|z|} \\ m + m' &= \begin{cases} s+1 & \text{if } h + \ell = |y| \\ s & \text{if } h + \ell = |z| + |y| \end{cases} \\ t + |z| &= (m + m')|z| + h + \ell. \end{aligned}$$

Observe also that $|y|$ is determined by $|z|$ ($|y| = t \pmod{|z|}$), that h is determined by ℓ and $|z|$, and that m' is determined by m, ℓ and $|z|$. Then

$$u \in \bigcup_{k=1}^{t-1}\bigcup_{\ell=0}^{k-1}\bigcup_{m=0}^{1+\lfloor\frac{t}{k}\rfloor}\bigcup_{z\in\mathcal{R}_k} X_{z,m,\ell}, \text{ where}$$
$$X_{z,\ell,m} = E((z^m\mathsf{pref}_\ell(z), \mathsf{suff}_h(z)z^{m'}), (1, n - m'|z| - h + 1))$$

and h and m' take the values imposed by those of $k = |z|$, ℓ and m. In particular, the probability P that a reduced word in $\mathcal{R}_n$ has overlapping occurrences of a cyclic factor of length t, with at least one of these occurrences straddling, satisfies

$$P \ \leq \ \sum_{k=1}^{t-1}\sum_{\ell=0}^{k-1}\sum_{m=0}^{1+\lfloor\frac{t}{k}\rfloor}\sum_{z\in\mathcal{R}_k} \mathbb{R}_n(X_{z,\ell,m}),$$

If $m \geq 1$, then

$$X_{z,\ell,m} = E((z, z^{m-1}\mathsf{pref}_\ell(z), \mathsf{suff}_h(z)z^{m'}), (1, |z| + 1, n - m'|z| - h + 1))$$

and a double application of Lemma 3.5 shows that

$$\mathbb{R}_n(X_{z,\ell,m}) \leq C^2\alpha^{m'|z|+h}\alpha^{(m-1)|z|+\ell}\mathbb{R}_n(E((z),(1))) = C^2\alpha^t\mathbb{R}_n(E((z),(1))).$$

Summing these over $z \in \mathcal{R}_k$ (with k, ℓ and m fixed, $m \geq 1$), we get

$$\sum_{z\in\mathcal{R}_k} \mathbb{R}_n(X_{z,\ell,m}) \leq \sum_{z\in\mathcal{R}_k} C^2\alpha^t\mathbb{R}_n(E((z),(1))) \leq C^2\alpha^t,$$

since $\mathcal{R}_n$ is partitioned by the $\mathbb{R}_n(E((z),(1)))$ ($z \in \mathcal{R}_k$).

If $m = 0$ and $h + \ell = |y|$, then $m'|z| = t + |z| - |y|$ and we note that

$$\begin{aligned} X_{z,\ell,0} &= E((\mathsf{pref}_\ell(z), \mathsf{suff}_h(z)z^{m'}), (1, n - t - |z| + \ell + 1)) \\ &\subseteq E((\mathsf{pref}_\ell(z), \mathsf{suff}_h(z), \mathsf{suff}_{|y|}(z)z^{m'-1}), (1, n - t - |z| + \ell + 1, n - t + 1)). \end{aligned}$$

By Lemma 3.5, we get

$$\mathbb{R}_n(X_{z,\ell,0}) \leq C\alpha^t \mathbb{R}_n(E((\mathsf{pref}_\ell(z), \mathsf{suff}_h(z)), (1, n - t - |z| + \ell + 1))).$$

Summing over all $z \in \mathcal{R}_k$ (k and ℓ fixed), we get

$$\begin{aligned} \sum_{z\in\mathcal{R}_k} \mathbb{R}_n(X_{z,\ell,0}) &\leq \sum_{z\in\mathcal{R}_k} C\alpha^t \mathbb{R}_n(E((\mathsf{pref}_\ell(z), \mathsf{suff}_h(z)), (1, n - t - k + \ell + 1))) \\ &\leq \sum_{z_1\in\mathcal{R}_\ell} \sum_{z_2\in\mathcal{R}_h} C\alpha^t \mathbb{R}_n(E((z_1, z_2), (1, n - t - k + \ell + 1))) \\ &\leq C\alpha^t, \end{aligned}$$

since $\mathcal{R}_n$ is partitioned by the $\mathbb{R}_n(E((z_1, z_2), (1, n - t - k + \ell + 1)))$ ($z_1 \in \mathcal{R}_\ell$, $z_2 \in \mathcal{R}_h$).

Finally, if $m = 0$ and $h + \ell = |z| + |y|$, then $m'|z| = t - |y|$. Therefore

$$\begin{aligned} X_{z,\ell,0} &= E((\mathsf{pref}_\ell(z), \mathsf{suff}_h(z)z^{m'}), (1, n - t - |z| + \ell + 1)) \\ &= E((\mathsf{pref}_\ell(z), \mathsf{pref}_{|z|-\ell}(\mathsf{suff}_h(z)), \mathsf{suff}_{|y|}(z)z^{m'}), (1, n - t - |z| + \ell + 1, n - t + 1)). \end{aligned}$$

By Lemma 3.5, this yields

$$\mathbb{R}_n(X_{z,\ell,0}) \leq C\alpha^t \mathbb{R}_n(E((\mathsf{pref}_\ell(z), \mathsf{pref}_{|z|-\ell}(\mathsf{suff}_h(z))), (1, n - t + |z| + \ell + 1))).$$

As in the previous case, summing over all $z \in \mathcal{R}_k$ (k and ℓ fixed) yields

$$\sum_{z\in\mathcal{R}_k} \mathbb{R}_n(X_{z,\ell,0}) \leq C\alpha^t.$$

Then we get the following upper bound for the probability P:

$$\begin{aligned} P &\leq \sum_{k=1}^{t-1} \sum_{\ell=0}^{k-1} \sum_{m=1}^{1+\lfloor \frac{t}{k} \rfloor} C^2\alpha^t + \sum_{k=1}^{t-1} \sum_{\ell=0}^{k-1} C\alpha^t \\ &\leq C^2 \frac{3}{2} t(t-1)\alpha^t + C\frac{1}{2} t(t-1)\alpha^t \\ &\leq \ 2C^2 t(t-1)\alpha^t. \end{aligned}$$

This concludes the proof. □

In order to extend the results of this section to cyclically reduced words, we need an additional hypothesis, essentially stating that the probability of cyclically reduced words does not vanish. In fact, we have the following general result.

LEMMA 3.13. *Let $(\mathbb{R}_n)_{n\geq 0}$ be a sequence of measures satisfying $\liminf \mathbb{R}_n(\mathcal{C}_n) = p > 0$. Let X be a subset of $\mathcal{R}$. Then for each $\delta > 1$ and for every large enough n, the probability $\mathbb{R}_n(X \mid \mathcal{C})$ that a cyclically reduced word of length n is in X is at most equal to $\frac{\delta}{p}\mathbb{R}_n(X)$. In particular, if X is exponentially (resp. super-polynomially, polynomially, simply) negligible, then so is $X \cap \mathcal{C}$ in $\mathcal{C}$.*

PROOF. By definition, $\mathbb{R}_n(X \mid \mathcal{C}) = \mathbb{R}_n(X \cap \mathcal{C} \mid \mathcal{C}) = \frac{\mathbb{R}_n(X\cap\mathcal{C})}{\mathbb{R}_n(\mathcal{C}_n)} \leq \frac{\delta}{p}\mathbb{R}_n(X)$, which concludes the proof. □

The following statement is an immediate consequence.

COROLLARY 3.14. *Let $(\mathbb{R}_n)_{n\geq 0}$ be a prefix-heavy sequence of parameters (C,α), with the property that $\liminf_n \mathbb{R}_n(\mathcal{C}_n) = p > 0$. Then for every $\delta > 1$ and every large enough n, the probability that a cyclically reduced word of length n has two non-overlapping occurrences of a cyclic factor of length t (resp. an occurrence of a cyclic factor of length t and its inverse, two overlapping occurrences of a cyclic factor of length t) is at most $\frac{\delta}{p}(Cn^2+C^2nt)\alpha^t$ (resp. $\frac{\delta}{p}(Cn^2+C^2nt)\alpha^t$, $\frac{3\delta}{p}C^2nt\alpha^t$).*

PROOF. Let X be the set of reduced words of length n with two non-overlapping occurrences of a cyclic factor of length t (resp. an occurrence of a cyclic factor of length t and its inverse, two overlapping occurrences of a cyclic factor of length t). It suffices to apply Lemma 3.13 to the set X, and to use the results of Lemmas 3.11 and 3.12. □

3.4. Measures on tuples of lengths and on tuples of words. For every positive integer k, let $\mathcal{T}_k$ denote the set of k-tuples of non-negative integers and $\mathcal{TW}_k$ denote the set of k-tuples of reduced words. Let also $\mathcal{T} = \bigcup_k \mathcal{T}_k$ and $\mathcal{TW} = \bigcup_k \mathcal{TW}_k$ be the sets of all tuples of non-negative integers, and of reduced words respectively.

For a given $\vec{h} = (h_1, \dots, h_k)$ of $\mathcal{TW}_k$, let $\|\vec{h}\|$ be the element of $\mathcal{T}_k$ given by

$$\|\vec{h}\| = (|h_1|, \dots, |h_k|)\,.$$

A *prefix-heavy sequence of measures on tuples of reduced words* is a sequence $(\mathbb{P}_n)_{n\geq 0}$ of measures on $\mathcal{TW}$ such that for every $\vec{h} = (h_1, \dots, h_k)$ of $\mathcal{TW}$,

$$\mathbb{P}_n(\vec{h}) = \mathbb{T}_n(\|\vec{h}\|) \prod_{i=1}^{k} \mathbb{R}_{|h_i|}(h_i),$$

where $(\mathbb{T}_n)_{n\geq 0}$ is a sequence of measures on $\mathcal{T}$ and $(\mathbb{R}_n)_{n\geq 0}$ is a prefix-heavy sequence of measures on $\mathcal{R}$. If $(\mathbb{R}_n)_{n\geq 0}$ is prefix-heavy with parameters (C,α), then we say that $(\mathbb{T}_n)_{n\geq 0}$ is prefix-heavy with parameters (C,α).

REMARK 3.15. In the definition above, to draw a tuple of words according to $\mathbb{P}_n$, one can first draw a tuple of lengths $(\ell_1, \dots, \ell_k)$ following $\mathbb{T}_n$, and then draw, independently for each coordinate, an element of $\mathcal{R}_{\ell_i}$ following $\mathbb{R}_{\ell_i}$.

EXAMPLE 3.16. Let $\nu(n)$ be an integer-valued function. The uniform distribution on the $\nu(n)$-tuples of reduced words of length exactly n is a prefix-heavy sequence of measures: one needs to take $\mathbb{T}_n$ to be the measure whose weight is entirely concentrated on the $\nu(n)$-tuple $(n, \dots, n)$ and $\mathbb{R}_n$ to be the uniform distribution on $\mathcal{R}_n$ (see Example 3.2).

The uniform distribution on the $\nu(n)$-tuples of reduced words of length at most n is also a prefix-heavy sequence of measures. Here the support of $\mathbb{T}_n$ must be restricted to the tuples $(x_1, \dots, x_{\nu(n)})$ such that $x_i \leq n$ for each i, with $\mathbb{T}_n(x_1, \dots, x_{\nu(n)}) = \prod_i \frac{|\mathcal{R}_{x_i}|}{|\mathcal{R}_{\leq n}|}$.

Both can be naturally adapted to handle the uniform distribution on the $\nu(n)$-tuples of cyclically reduced words of length exactly (resp. at most) n.

For appropriate functions $\nu(n)$, we retrieve the few-generator and the density models discussed in Section 2.2. We will see a more general class of examples in Section 4.

3.5. General statements. If $\vec{x} \in \mathcal{T}$, we denote by $\mathsf{max}(\vec{x})$ and $\mathsf{min}(\vec{x})$ the maximum and minimum element of $\vec{x}$. We also denote by $\mathsf{size}(\vec{x})$ the integer k such that $\vec{x} \in \mathcal{T}_k$.

The statistics min, max, and size are extended to tuples of words by setting $\mathsf{min}(\vec{h}) = \mathsf{min}(\|\vec{h}\|)$, $\mathsf{max}(\vec{h}) = \mathsf{max}(\|\vec{h}\|)$ and $\mathsf{size}(\vec{h}) = \mathsf{size}(\|\vec{h}\|)$. In the sequel we consider sequences of probability spaces on $\mathcal{TW}$ and min, max, and size are seen as random variables.

The following statements give general sufficient conditions for a tuple to generically have the central tree property, generate a malnormal subgroup, or satisfy a small cancellation property.

PROPOSITION 3.17. *Let $(\mathbb{P}_n)_{n\geq 0}$ be a prefix-heavy sequence of measures on tuples of reduced words of parameters (C, α). Let $f\colon \mathbb{N} \to \mathbb{N}$ such that $f(\ell) \leq \frac{\ell}{2}$ for each ℓ. If there exists a sequence $(\eta_n)_{n\geq 0}$ of positive real numbers such that*

$$\lim_{n\to\infty} \mathbb{P}_n\left(\mathsf{size}^2\, \alpha^{f(\mathsf{min})} > \eta_n\right) = 0 \quad \textit{and} \quad \lim_{n\to\infty} \eta_n = 0, \tag{1}$$

then a random tuple of words generically satisfies $\mathsf{lcp}(\vec{h}) < f(\mathsf{min}(\vec{h}))$.

If the limits in Equation (1) converge polynomially (resp. super-polynomially, exponentially) fast, then $\mathsf{lcp}(\vec{h}) < f(\mathsf{min}(\vec{h}))$ polynomially (resp. super-polynomially, exponentially) generically.

PROOF. The set of all tuples $\vec{h}$ that fail to satisfy the inequality $\mathsf{lcp}(\vec{h}) < f(\mathsf{min}(\vec{h}))$ is the union $\mathcal{G}_1 \cup \mathcal{G}_2$ of the two following sets:

- the set $\mathcal{G}_1$ of all tuples $\vec{h} = (h_1, \dots, h_k)$ such that for some $1 \leq i < j \leq k$, a word of length $f(\mathsf{min}(\vec{h}))$ occurs as a prefix of h_i or h_i^{-1}, and also of h_j or h_j^{-1},
- the set $\mathcal{G}_2$ of all tuples $\vec{h} = (h_1, \dots, h_k)$ such that for some $1 \leq i \leq k$, h_i and h_i^{-1} have a common prefix of length $f(\mathsf{min}(\vec{h}))$,

and we only need to prove that $\lim_n \mathbb{P}_n(\mathcal{G}_1) = \lim_n \mathbb{P}_n(\mathcal{G}_2) = 0$.

Let k, ℓ be positive integers and let $X_{k,\ell}$ be the set of tuples $\vec{h} \in \mathcal{TW}_k$ such that $\mathsf{min}(\vec{h}) = \ell$. If $\vec{h} \in X_{k,\ell}$ and $1 \leq i < j \leq k$, then the probability that h_i and h_j have the same prefix of length $t = f(\ell)$ is

$$\sum_{w\in\mathcal{R}_t} \mathbb{R}_{|h_i|}(\mathcal{P}(w))\mathbb{R}_{|h_j|}(\mathcal{P}(w)) \;\leq\; C\alpha^t \sum_{w\in\mathcal{R}_t} \mathbb{R}_{|h_j|}(\mathcal{P}(w)) \;\leq\; C\alpha^t.$$

Then we have $\mathbb{P}_n(\mathcal{G}_1 \mid X_{k,\ell}) \leq 4k^2 C\alpha^{f(\ell)}$, or rather $\mathbb{P}_n(\mathcal{G}_1 \mid X_{k,\ell}) \leq \min(1, 4k^2 C\alpha^{f(\ell)})$, where the factor k^2 corresponds to the choice of i and j and the factor 4 corresponds to the possibilities that h_i or h_i^{-1}, and h_j or h_j^{-1} have a common prefix of length $f(\ell)$. Therefore we have $\mathbb{P}_n(\mathcal{G}_1 \cap X_{k,\ell}) \leq \min(1, 4k^2 C\alpha^{f(\ell)})\ \mathbb{P}_n(X_{k,\ell})$

We can split the set of pairs (k, ℓ) into those pairs such that $k^2\alpha^{f(\ell)} > \eta_n$ and the others, for which $k^2\alpha^{f(\ell)} \leq \eta_n$. Then we have

$$\mathbb{P}_n(\mathcal{G}_1) \;=\; \sum_{k,\ell} \mathbb{P}_n(\mathcal{G}_1 \cap X_{k,\ell}) \;\leq\; \mathbb{P}_n(\mathsf{size}^2\, \alpha^{f(\mathsf{min})} > \eta_n) + 4C\ \eta_n,$$

which tends to 0 under the hypothesis in Equation (1).

Similarly, if $\vec{h} \in X_{k,\ell}$ and $i \leq k$, the probability that h_i and h_i^{-1} have a common prefix of length $f(\ell)$ is at most $C\alpha^{f(\ell)}$ by Corollary 3.8. It follows that $\mathbb{P}_n(\mathcal{G}_2 \mid X_{k,\ell}) \leq \min(1, kC\alpha^{f(\ell)})$, and $\mathbb{P}_n(\mathcal{G}_2 \cap X_{k,\ell}) \leq \min(1, kC\alpha^{f(\ell)})\ \mathbb{P}_n(X_{k,\ell})$.

Splitting the set of pairs (k,ℓ) into those pairs such that $k\alpha^{f(\ell)} > \eta_n$ and those for which $k\alpha^{f(\ell)} \leq \eta_n$, yields

$$\mathbb{P}_n(\mathcal{G}_2) \;=\; \sum_{k,\ell} \mathbb{P}_n(\mathcal{G}_2 \cap X_{k,\ell}) \;\leq\; \mathbb{P}_n(\mathsf{size}\,\alpha^{f(\mathsf{min})} > \eta_n) + C\ \eta_n.$$

Now $\mathsf{size}\,\alpha^{f(\mathsf{min})} < \mathsf{size}^2\,\alpha^{f(\mathsf{min})}$, so $\mathbb{P}_n(\mathsf{size}\,\alpha^{f(\mathsf{min})} > \eta_n) \leq \mathbb{P}_n(\mathsf{size}^2\,\alpha^{f(\mathsf{min})} > \eta_n)$. It follows that $\lim_n \mathbb{P}_n(\mathsf{size}\,\alpha^{f(\mathsf{min})} > \eta_n) = 0$, and hence $\lim_n \mathbb{P}_n(\mathcal{G}_2) = 0$, which concludes the proof. □

THEOREM 3.18 (Central tree property). *Let $(\mathbb{P}_n)_{n\geq 0}$ be a prefix-heavy sequence of measures on tuples of reduced words of parameters (C,α). If there exists a sequence $(\eta_n)_{n\geq 0}$ of positive real numbers such that*

$$\lim_{n\to\infty} \mathbb{P}_n\left(\mathit{size}^2\,\alpha^{\frac{\mathit{min}}{2}} > \eta_n\right) = 0 \quad \textit{and} \quad \lim_{n\to\infty} \eta_n = 0, \tag{2}$$

then a random tuple of words generically has the central tree property. In particular, such a tuple is a basis of the subgroup it generates.

If the limits in Equation (2) converge polynomially (resp. super-polynomially, exponentially) fast, then the central tree property holds polynomially (resp. super-polynomially, exponentially) generically.

PROOF. By definition, a tuple $\vec{h} \in \mathcal{TW}$ satisfies the central tree property if $\mathsf{lcp}(\vec{h}) < \frac{\mathsf{min}(\vec{h})}{2}$, so the theorem is a direct application of Proposition 3.17 to the function $f(\ell) = \frac{\ell}{2}$, and of Proposition 1.2. □

THEOREM 3.19 (Malnormality). *Let $(\mathbb{P}_n)_{n\geq 0}$ be a prefix-heavy sequence of measures on tuples of reduced words of parameters (C,α). If there exists a sequence $(\eta_n)_{n\geq 0}$ of positive real numbers such that*

$$\lim_{n\to\infty} \mathbb{P}_n\left(\mathit{size}^2\,\mathit{max}^2\,\alpha^{\frac{\mathit{min}}{8}} > \eta_n\right) = 0 \quad \textit{and} \quad \lim_{n\to\infty} \eta_n = 0, \tag{3}$$

then a random tuple of words generically generates a malnormal subgroup.

If the limits in Equation (3) converge polynomially (resp. super-polynomially, exponentially) fast, then malnormality holds polynomially (resp. super-polynomially, exponentially) generically.

PROOF. By Proposition 1.3, a sufficient condition for a tuple $\vec{h} \in \mathcal{TW}$ to generate a malnormal subgroup is to have $\mathsf{lcp}(\vec{h}) < \frac{1}{3}\,\mathsf{min}(\vec{h})$, and to not have two occurrences of a word of length $\frac{1}{2}(\mathsf{min}(\vec{h}) - 3\,\mathsf{lcp}(\vec{h}))$ as a factor of a word in $\vec{h}^{\pm}$. This condition is satisfied in particular if $\mathsf{lcp}(\vec{h}) < \frac{1}{4}\,\mathsf{min}(\vec{h})$ and no word of length $\frac{1}{8}\,\mathsf{min}(\vec{h})$ has two occurrences as a factor of a word in $\vec{h}^{\pm}$.

Therefore the set of all tuples $\vec{h}$ that generate a non malnormal subgroup is contained in the union $\mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3 \cup \mathcal{G}_4$ of the following sets:

- the set $\mathcal{G}_1$ of all tuples $\vec{h} = (h_1, \dots, h_k)$ such that $\mathsf{lcp}(\vec{h}) \geq \frac{1}{4}\,\mathsf{min}(\vec{h})$,
- the set $\mathcal{G}_2$ of all tuples $\vec{h} = (h_1, \dots, h_k)$ such that for some $1 \leq i < j \leq k$, a word of length $\frac{1}{8}\,\mathsf{min}(\vec{h})$ occurs as a factor of h_i, and also of h_j or h_j^{-1},
- the set $\mathcal{G}_3$ of all tuples $\vec{h} = (h_1, \dots, h_k)$ such that for some $1 \leq i \leq k$, h_i and h_i^{-1} have a common factor of length $\frac{1}{8}\,\mathsf{min}(\vec{h})$,
- the set $\mathcal{G}_4$ of all tuples $\vec{h} = (h_1, \dots, h_k)$ such that for some $1 \leq i \leq k$, h_i has at least two occurrences of a factor of length $\frac{1}{8}\,\mathsf{min}(\vec{h})$,

and we want to verify that $\mathbb{P}_n(\mathcal{G}_1)$, $\mathbb{P}_n(\mathcal{G}_2)$, $\mathbb{P}_n(\mathcal{G}_3)$ and $\mathbb{P}_n(\mathcal{G}_4)$ all tend to 0 when n tends to infinity.

By Proposition 3.17, the set $\mathcal{G}_1$ is negligible as soon as $\lim_n \mathbb{P}_n(\mathsf{size}\,\alpha^{\frac{\mathsf{min}}{4}} > \eta_n) = 0$. This is true under the hypothesis in Equation (3) since $\mathsf{size}\,\alpha^{\frac{\mathsf{min}}{4}} < \mathsf{size}^2\,\mathsf{max}^2\,\alpha^{\frac{\mathsf{min}}{8}}$, and hence $\mathbb{P}_n(\mathsf{size}\,\alpha^{\frac{\mathsf{min}}{4}} > \eta_n) \leq \mathbb{P}_n(\mathsf{size}^2\,\mathsf{max}^2\,\alpha^{\frac{\mathsf{min}}{8}} > \eta_n)$.

Let now $X_{k,\ell,M}$ be the set of tuples $\vec{h} \in X_{k,\ell}$ such that $\mathsf{max}(\vec{h}) = M$. Let $1 \leq i < j \leq k$ and $\vec{h} \in X_{k,\ell,M}$. By Corollary 3.6, the probability that h_j has a given factor v of length $\frac{\ell}{8}$ is at most equal to $CM\alpha^{\frac{\ell}{8}}$. Summing this probability over all words v which occur as a factor of h_i (at most $|h_i| \leq M$ such words), it follows that the probability that h_i and h_j have a common factor of length $t = \frac{\ell}{8}$ is at most equal to $CM^2\alpha^{\frac{\ell}{8}}$. Summing now over the possible values of i and j, we find that $\mathbb{P}_n(\mathcal{G}_2 \cap X_{k,\ell,M}) \leq \min(1, k^2CM^2\alpha^{\frac{\ell}{8}})\ \mathbb{P}_n(X_{k,\ell,M})$ and therefore, as above

$$\mathbb{P}_n(\mathcal{G}_2) \ \leq\ \mathbb{P}_n(\mathsf{size}^2\,\mathsf{max}^2\,\alpha^{\frac{\mathsf{min}}{8}} > \eta_n) + C\ \eta_n.$$

It follows from Equation (3) that $\mathcal{G}_2$ is negligible.

By Lemma 3.7, the probability that h_i and h_i^{-1} have a common factor of length $\frac{\ell}{8}$ is at most $CM^2\alpha^{\frac{\ell}{8}}$. Summing over all choices of i, we find that

$$\mathbb{P}_n(\mathcal{G}_3) \ \leq\ \mathbb{P}_n(\mathsf{size}\,\mathsf{max}^2\,\alpha^{\frac{\mathsf{min}}{8}} > \eta_n) + C\ \eta_n.$$

Since $\mathsf{size}\,\mathsf{max}^2\,\alpha^{\frac{\mathsf{min}}{8}} < \mathsf{size}^2\,\mathsf{max}^2\,\alpha^{\frac{\mathsf{min}}{8}}$, we conclude that $\mathcal{G}_3$ is negligible.

Finally, we have $\mathbb{P}_n(\mathcal{G}_4) \leq \frac{C}{8}\,\mathsf{size}\,\mathsf{max}\,\mathsf{min}\,\alpha^{\frac{\mathsf{min}}{8}}$ by Lemma 3.9, and hence

$$\mathbb{P}_n(\mathcal{G}_4) \ \leq\ \mathbb{P}_n(\mathsf{size}\,\mathsf{max}\,\mathsf{min}\,\alpha^{\frac{\mathsf{min}}{8}} > \eta_n) + \frac{C}{8}\ \eta_n.$$

Since $\mathsf{size}\,\mathsf{max}\,\mathsf{min}\,\alpha^{\frac{\mathsf{min}}{8}} < \mathsf{size}^2\,\mathsf{max}^2\,\alpha^{\frac{\mathsf{min}}{8}}$, it follows as above that the set $\mathcal{G}_4$ is negligible. □

THEOREM 3.20 (Small cancellations property). *Let $(\mathbb{P}_n)_{n\geq 0}$ be a prefix-heavy sequence of measures on tuples of reduced words of parameters (C, α), such that $\liminf_n \mathbb{R}_n(\mathcal{C}_n) = p > 0$. For any $\lambda \in (0, \frac{1}{2})$, if there exists a sequence $(\eta_n)_{n\geq 0}$ of positive real numbers such that*

$$\lim_{n\to\infty} \mathbb{P}_n\left(\mathsf{size}^2\ \mathsf{max}^2\,\alpha^{\lambda\,\mathsf{min}} > \eta_n\right) = 0 \quad \text{and} \quad \lim_{n\to\infty} \eta_n = 0, \tag{4}$$

then the property $C'(\lambda)$ generically holds.

If the limits in Equation (4) converge polynomially (resp. super-polynomially, exponentially) fast, then Property $C'(\lambda)$ holds polynomially (resp. super-polynomially, exponentially) generically.

PROOF. A sufficient condition for a tuple of cyclically reduced words $\vec{h}$ to satisfy $C'(\lambda)$ is for every piece in $\vec{h}$ to have length less than $\lambda\,\mathsf{min}(\vec{h})$. Then the set $\mathcal{G}$ of tuples that fail to satisfy $C'(\lambda)$ is contained in the union $\mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3 \cup \mathcal{G}_4$ of the following sets:

- the set $\mathcal{G}_1$ of all tuples of cyclically reduced words $\vec{h} = (h_1, \ldots, h_k)$ such that for some $1 \leq i < j \leq k$, a word of length $\lambda\,\mathsf{min}(\vec{h})$ occurs as a factor of h_i, and also of h_j or h_j^{-1},
- the set $\mathcal{G}_2$ of all tuples of cyclically reduced words $\vec{h} = (h_1, \ldots, h_k)$ such that for some $1 \leq i \leq k$, h_i has two non-overlapping occurrences of a factor of length $\lambda\,\mathsf{min}(\vec{h})$,

- the set $\mathcal{G}_3$ of all tuples of cyclically reduced words $\vec{h} = (h_1, \dots, h_k)$ such that for some $1 \leq i \leq k$, h_i has non-overlapping occurrences of a factor of length $\lambda \, \mathsf{min}(\vec{h})$ and its inverse,
- the set $\mathcal{G}_4$ of all tuples of cyclically reduced words $\vec{h} = (h_1, \dots, h_k)$ such that for some $1 \leq i \leq k$, h_i has overlapping occurrences of a factor of length $\lambda \, \mathsf{min}(\vec{h})$,

and we want to verify that $\mathbb{P}_n(\mathcal{G}_1)$, $\mathbb{P}_n(\mathcal{G}_2)$, $\mathbb{P}_n(\mathcal{G}_3)$ and $\mathbb{P}_n(\mathcal{G}_4)$ all tend to 0 when n tends to infinity.

As in the proof of Theorem 3.19, we find that the probability that a tuple of reduced words $\vec{h}$ is such that a word of length $\lambda \, \mathsf{min}(\vec{h})$ occurs as a factor of h_i, and also of h_j or h_j^{-1}, for some $i < j$ is at most $\mathbb{P}_n(\mathsf{size}^2 \, \mathsf{max}^2 \, \alpha^{\lambda \, \mathsf{min}} > \eta_n) + C \; \eta_n$. Reasoning as in the proof of Corollary 3.14, it follows that, for every $\delta > 1$,

$$\mathbb{P}_n(\mathcal{G}_1) \leq \frac{\delta}{p} \left(\mathbb{P}_n(\mathsf{size}^2 \, \mathsf{max}^2 \, \alpha^{\lambda \, \mathsf{min}} > \eta_n) + C \; \eta_n \right),$$

and it follows from Equation (4) that $\mathcal{G}_1$ is negligible.

Now using Corollary 3.14, we show that

$$\mathbb{P}_n(\mathcal{G}_2), \mathbb{P}_n(\mathcal{G}_3) \leq \frac{\delta}{p} \left(\mathbb{P}_n(\mathsf{size}(\mathsf{max}^2 + \mathsf{max} \, \mathsf{min}) \alpha^{\lambda \, \mathsf{min}} > \eta_n) + C^2 \eta_n \right),$$

$$\mathbb{P}_n(\mathcal{G}_4) \leq \frac{\delta}{p} \left(\mathbb{P}_n(\mathsf{size}(\mathsf{max} \, \mathsf{min} + \mathsf{min}^2) \alpha^{\lambda \, \mathsf{min}} > \eta_n) + 2C^2 \eta_n \right).$$

Since $\mathsf{size} \, \mathsf{max}^2$, $\mathsf{size} \, \mathsf{max} \, \mathsf{min}$ and $\mathsf{size} \, \mathsf{min}^2$ are less than $\mathsf{size}^2 \, \mathsf{max}^2$, the hypothesis in Equation (4) shows that $\mathcal{G}_2$, $\mathcal{G}_3$ and $\mathcal{G}_4$ are negligible, and this concludes the proof. □

3.6. Applications to the uniform distribution case. The few-generator model and the density model, based on the uniform distribution on reduced words of a given length and discussed in Section 2.2, are both instances of a prefix-heavy sequence of measures on tuples, for which the parameter α is $\alpha = \frac{1}{2r-1}$, see Examples 3.2 and 3.16. In this section, the measure $\mathbb{R}_n$ is the uniform distribution on $\mathcal{R}_n$.

The results of Section 3.5 above allow us to retrieve many of the results in Section 2.2 — typically the results on the small cancellation property $C'(\lambda)$ up to density $\frac{\lambda}{2}$, whether one considers tuples of cyclically reduced words of length n or of length at most n —, and to expand them. In particular, we show that the results on the central tree property and malnormality in the few-generator model can be extended to the density model, and that we have a phase transition theorem for the central tree property (at density $\frac{1}{4}$).

Small cancellation properties Let $0 < d < 1$. In the density model, at density d, we choose uniformly at random a $\nu(n)$-tuple of cyclically reduced words of length n, with $\nu(n) = |\mathcal{C}_n|^d$. In particular, for every tuple $\vec{h}$ of that sort, we have $\mathsf{size}(\vec{h}) = \nu(n)$ and $\mathsf{max}(\vec{h}) = \mathsf{min}(\vec{h}) = n$.

Let $0 < \lambda < \frac{1}{2}$ and for each n, let

$$\eta_n \;=\; \left(\frac{2r}{2r-1} \right)^{2d} n^2 \; (2r-1)^{-(\lambda - 2d)n} \;+\; \left(\frac{2r}{2r-1} \right)^{d} n^2 \; (2r-1)^{-(\lambda - d)n}.$$

Note that $|\mathcal{C}_n| < |\mathcal{R}_n| = \frac{2r}{2r-1}(2r-1)^n$. Therefore $\mathsf{size}^2\,\mathsf{max}^2\,\alpha^{\lambda\,\mathsf{min}} < \eta_n$ with probability 1. Now observe that η_n converges exponentially fast to 0 when $d < \frac{\lambda}{2}$. In view of Theorem 3.20, this provides a proof of part of Theorem 2.4 (2), namely, of the fact that, at density less than $\frac{\lambda}{2}$, Property $C'(\lambda)$ holds exponentially generically.

It is unclear whether the more difficult property, that hyperbolicity holds generically at density less than $\frac{1}{2}$, can be established with the same very general tools.

Observe that the set $\mathcal{R}_{\leq n}$ of reduced words of length at most n has cardinality $1 + \sum_{i=1}^{n} |\mathcal{R}_n| = \frac{r}{r-1}(2r-1)^n - \frac{1}{r-1}$. By the same reasoning as above, at density less than $\frac{\lambda}{2}$, a tuple of cyclically reduced words of length at most n exponentially generically has Property $C'(\lambda)$.

Properties of subgroups We now return to tuples of reduced words like in the few-generator model, but with a density type assumption on the size of the tuples. For $0 < d < 1$, we consider $|\mathcal{R}_{\leq n}|^d$-tuples of reduced words of length at most n, and the asymptotic properties of the subgroups generated by these tuples. For such tuples $\vec{h}$, we have $\mathsf{size}(\vec{h}) \leq \left(\frac{r}{r-1}\right)^d (2r-1)^{dn}$ and $\mathsf{max}(\vec{h}) = n$.

In addition, for every $0 < \mu < 1$, Proposition 2.1 shows that $\mathsf{min}(\vec{h}) > \mu n$, exponentially generically.

We first establish the central tree property.

Proposition 3.21. *Let $0 < d < \frac{1}{4}$. At density d, a tuple of reduced words of length at most n chosen uniformly at random, exponentially generically has the central tree property, and in particular it is a basis of the subgroup it generates.*

If $d > \frac{1}{4}$, then at density d the central tree property exponentially generically does not hold.

Proof. For a fixed $\mu < 1$, the following inequality holds exponentially generically:

$$\mathsf{size}^2\,\alpha^{\frac{\mathsf{min}}{2}} \leq \left(\frac{r}{r-1}\right)^{2d} (2r-1)^{-(\frac{\mu}{2}-2d)n}.$$

At every density $d < \frac{1}{4}$, one can choose $\mu < 1$ such that $\frac{\mu}{2} - 2d > 0$ (say, $\mu = \frac{1+4d}{2}$). For such a value of μ, $\eta_n = \left(\frac{r}{r-1}\right)^{2d} (2r-1)^{-(\frac{\mu}{2}-2d)n}$ converges exponentially fast to 0 and, in view of Theorem 3.18, this proves the first part of the proposition.

If $d > \frac{1}{4}$, let d' be such that $\frac{1}{4} < d' < \min(\frac{1}{2}, d)$. By the classical Birthday Paradox[3], exponentially generically two words of the tuple share a prefix of length $2d'n$. This prove the second part of the proposition. □

Along the same lines, we also prove the following result.

Proposition 3.22. *Let $0 < d < \frac{1}{16}$. At density d, a tuple of reduced words of length at most n chosen uniformly at random, exponentially generically generates a malnormal subgroup.*

Proof. For a fixed $\mu < 1$, we have

$$\mathsf{size}^2\,\mathsf{max}^2\,\alpha^{\frac{\mathsf{min}}{8}} \leq \left(\frac{r}{r-1}\right)^{2d} n^2 (2r-1)^{-(\frac{\mu}{8}-2d)n},$$

[3]If E is a set of size M and x is a uniform random tuple of E^m, the probability that the coordinates of x are pairwise distinct is $(1-\frac{1}{M})(1-\frac{2}{M})\cdots(1-\frac{m-1}{M})$, which is at most $\exp(-\frac{m(m-1)}{2M})$ by direct calculations.

exponentially generically.

If $d < \frac{1}{16}$, one can choose $\mu < 1$ such that $\frac{\mu}{8} - 2d > 0$ (say, $\mu = \frac{1+16d}{2}$), and we conclude as above, letting

$$\eta_n = \left(\frac{r}{r-1}\right)^{2d} n^2 (2r-1)^{-(\frac{\mu}{8}-2d)n}$$

and using Theorem 3.19. □

REMARK 3.23. Propositions 3.21 and 3.22 above generalize Corollary 2.2 (1) and (2), from the few generator case to an exponential number of generators — up to density $\frac{1}{4}$ and $\frac{1}{16}$, respectively (see Proposition 1.5).

Proposition 3.21 can actually be radically refined if the tuples have less than exponential size and if we drop the requirement of exponential genericity.

PROPOSITION 3.24. *Let f be an unbounded non-decreasing integer function. Let $k > 1$ be a fixed integer. Then a k-tuple $\vec{h}$ of reduced words of length at most n chosen uniformly at random, generically has the central tree property, with $\mathsf{lcp}(\vec{h}) \leq f(n)$.*

Let $c, c' > 0$ such that $c' \log(2r-1) > 2c$. Then an n^c-tuple $\vec{h}$ of reduced words of length at most n chosen uniformly at random, generically has the central tree property, with $\mathsf{lcp}(\vec{h}) \leq c' \log n$.

PROOF. If k is a fixed integer, then as in the proof of Proposition 3.21, we find that, for each $\mu < 1$, $\mathsf{size}^2\, \alpha^{f(\mathsf{min})}$ is generically less than or equal to $\eta_n = k^2(2r-1)^{-f(\mu n)}$, which tends to 0. This concludes the proof on the size of the central tree of random k-tuples by Proposition 3.17.

If we now consider n^c-tuples, we find that, for each $\mu < 1$, $\mathsf{size}^2\, \alpha^{c' \log(\mu n))}$ is generically less than or equal to $\eta_n = n^{2c}(2r-1)^{-c' \log n} = n^{-(c' \log(2r-1)-2c)}$, which tends to 0. By Proposition 3.17 again, this concludes the proof. □

4. Markovian automata

We now switch from the very general settings of the previous section to a specific and computable way to define prefix-heavy sequences of measures on reduced words.

We introduce Markovian automata (Section 4.1) which determine prefix-heavy sequences of measures under a simple and natural non-triviality assumption. These automata are a form of hidden Markov chain, and when they have a classical ergodicity property, then cyclically reduced words have asymptotically positive density. We are then able to generalize the results of Section 3.6 about central tree property and malnormality.

In the last part of the section, we give a generalization of Theorem 2.4 (2) and (3) on small cancellation and the degeneracy of a finite presentation.

4.1. Definition and examples.

A *Markovian automaton*[4] $\mathcal{A}$ consists of

- a deterministic transition system $(Q, \cdot)$ on alphabet X, where Q is a finite non-empty set called the *state set*, and for each $q \in Q$, $x \in X$, $q \cdot x \in Q$ or $q \cdot x$ is undefined;

[4]This notion is different from the two notions of probabilistic automata, introduced by Rabin [**26**] and Segala and Lynch [**28**], respectively.

- an initial probability vector $\gamma_0 \in [0,1]^Q$, that is, a positive vector such that $\sum_{q\in Q} \gamma_0(q) = 1$;
- for each $p \in Q$, a probability vector $(\gamma(p,x))_{x\in X} \in [0,1]^X$, such that $\gamma(p,x) = 0$ if and only if $p \cdot x$ is undefined.

If $u = x_0 \cdots x_n \in X^*$ $(n \geq 0)$, we write $\gamma(q,u) = \gamma(q,x_0)\gamma(q\cdot x_0, x_1)\cdots\gamma(q\cdot(x_0\cdots x_{n-1}), x_n)$. We let $\gamma(q,u) = 1$ if u is the empty word. We also write $\gamma_0(u) = \sum_{q\in Q}\gamma_0(q)\gamma(q,u)$.

Markovian automata are very similar to hidden Markov chain models, except that symbols are output on transitions instead of on states. We will discuss this further in Section 4.2 below. Markovian automata can be considered as more intuitive since sets of words (languages) are naturally described by automata.

We observe that, for each $n \geq 0$, $\sum_{|u|=n}\gamma(u) = 1$. Thus γ determines a probability measure $\mathbb{R}_n$ on the set of elements of X^* of length n: if $|u| = n$, then $\mathbb{R}_n(u) = \gamma(u)$.

In the sequel, we consider only Markovian automata on alphabet $\tilde{A}$, where only reduced words have non-zero probability. More precisely, the *support* of a Markovian automaton $\mathcal{A}$ is the set of words that can be read in $\mathcal{A}$, starting from a state q such that $\gamma_0(q) \neq 0$, that is, the set of all words u such that $\gamma(u) \neq 0$: we assume that our Markovian automata are such that their support is contained in $\mathcal{R}$.

EXAMPLE 4.1. *Uniform distribution on reduced words of length n.* It is immediately verified that the following Markovian automaton yields the uniform distribution on reduced words of each possible length. The state set is $Q = \tilde{A}$. For each $a \in \tilde{A}$, there is an a-labeled transition from every state except a^{-1}, ending in state a. All these transitions have the same probability, namely $\frac{1}{2r-1}$, and the initial probability vector is uniform as well, with each coordinate equal to $\frac{1}{2r}$.

One can also tweak these probabilities, to favor certain letters over others, or to favor positive letters (the letters in A) over negative letters.

EXAMPLE 4.2. *Distributions on rational subsets of $F(A)$.* The support of a Markovian automaton $\mathcal{A}$ is always rational and closed under taking prefixes, but it does not have to be equal to the set of all reduced words. We can consider a rational subset L of $F(A)$, or rather a deterministic transition system reading only reduced words, and impose probabilistic weights on its transitions to form a Markovian automaton. The resulting distribution gives non-zero weights only to prefixes of elements of L.

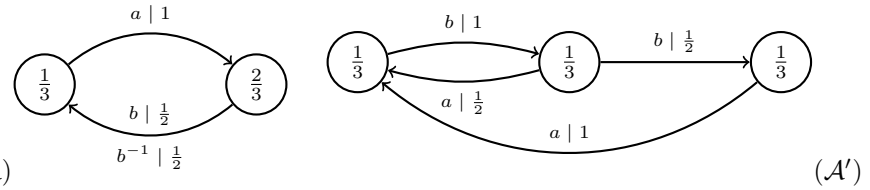

FIGURE 3. Markovian automata $\mathcal{A}$ and $\mathcal{A}'$.

Figure 3 represents two such automata (transitions are labeled by a letter and a probability, and each state is decorated with the corresponding initial probability), which are related with the modular group, $PSL(2,\mathbb{Z}) = \langle a, b \mid a^2, b^3\rangle$.

The support of the distribution defined by automaton $\mathcal{A}$ is the set of words over alphabet $\{a, b, b^{-1}\}$ without occurrences of the factors a^2, b^2, $(b^{-1})^2$, bb^{-1} and $b^{-1}b$, and the support of the distribution defined by $\mathcal{A}'$ consists of the words on alphabet $\{a, b\}$, without occurrences of a^2 or b^3. Both are regular sets of unique representatives of the elements of $PSL(2, \mathbb{Z})$: the first is the set of geodesics of $PSL(2, \mathbb{Z})$, and also the set of Dehn-reduced words with respect to the given presentation of that group; the second is a set of quasi-geodesics of $PSL(2, \mathbb{Z})$. Notice that the distribution produced by $\mathcal{A}'$ is not uniform on words of length n of its support.

Example 4.1 shows that the sequence $(\mathbb{R}_n)_n$ of uniform measures on reduced words, discussed in Sections 2.2 and 3.6 can be specified by a Markovian automaton. We also know that this sequence is prefix-heavy (Example 3.2). This is a general fact, under mild assumptions on the Markovian automaton.

PROPOSITION 4.3. *Let $\mathcal{A}$ be a Markovian automaton and let $(\mathbb{R}_n)_n$ the sequence of probability measures it determines. If $\mathcal{A}$ does not have a cycle with probability 1, then $(\mathbb{R}_n)_n$ is a prefix-heavy sequence of measures, with computable parameters (C, α).*

PROOF. Let ℓ be the maximum length of an elementary cycle (one that does not visit twice the same state) and let δ be the maximum value of $\gamma(q, \kappa)$ where κ is an elementary cycle at state q. Under our hypothesis, $\delta < 1$.

Every cycle κ can be represented as a composition of at least $|\kappa|/\ell$ elementary cycles (here, the composition takes the form of a sequence of insertions of a cycle in another). Consequently $\gamma(q, \kappa) \leq \delta^{\frac{|\kappa|}{\ell}}$. Finally, every path can be seen as a product of cycles and at most $|Q|$ individual edges. So, if u is a word and $q \in Q$, then $\gamma(q, u) \leq \delta^{\frac{|u|-|Q|}{\ell}}$, that is $\gamma(q, u) \leq C\alpha^{|u|}$ where $C = \delta^{\frac{-|Q|}{\ell}}$ and $\alpha = \delta^{\frac{1}{\ell}}$.

Let u, v be reduced words such that uv is reduced and let $n \geq |uv|$. We have

$$\begin{aligned}\mathbb{R}_n(\mathcal{P}(uv)) = \gamma_0(uv) &= \sum_{p \in Q} \gamma_0(p)\gamma(p, u)\gamma(p \cdot u, v) \\ &\leq \left(\sum_{p \in Q} \gamma_0(p)\gamma(p, u) \right) C\alpha^{|v|} \\ &= \gamma_0(u)\ C\alpha^{|v|} = \mathbb{R}_n(\mathcal{P}(u))\ C\alpha^{|v|},\end{aligned}$$

and hence $\mathbb{R}_n(\mathcal{P}(uv) \mid \mathcal{P}(u)) \leq C\alpha^{|v|}$, which concludes the proof. □

REMARK 4.4. The parameters C and α described in the proof of Proposition 4.3 may be far from optimal. If $\beta < 1$ is a uniform bound on the probabilities of the transitions of $\mathcal{A}$, then $\gamma_0(v), \gamma(q, v) \leq \beta^{|v|}$ for each word v, and the computation in the proof above shows that $\mathbb{R}_n(\mathcal{P}(uv) \mid \mathcal{P}(u)) \leq \beta^{|v|}$. We will see in Section 4.2 that we can be more precise under additional hypotheses.

Now let $\mathcal{A}$ be a Markovian automaton without a probability 1 cycle, such that the sequence of probability measures it induces is prefix-heavy with parameters (C, α). If $0 < d < 1$, we say that a tuple $\vec{h}$ of reduced words of length at most (resp. exactly) n is chosen at random according to $\mathcal{A}$, *at α-density d* if $\vec{h}$ consists of α^{-dn} words. Observe that this generalizes the concept discussed in Section 2.2.2 and 3.6.

With the same proofs as in Section 3.6, we have the following generalization of Propositions 3.21 and 3.22 related to central tree property and malnormality.

COROLLARY 4.5. *Let $\mathcal{A}$ be a Markovian automaton without a probability 1 cycle, such that the induced sequence of probability measures is prefix-heavy with parameters (C, α). Then a tuple of reduced words of length at most n chosen at random according to $\mathcal{A}$, at α-density $d < \frac{1}{4}$, exponentially generically has the central tree property.*

At α-density $d < \frac{1}{16}$, it exponentially generically generates a malnormal subgroup.

4.2. Irreducible Markovian automata and coincidence probability. An (n, n)-matrix M is said to be *irreducible* if it has non-negative coefficients and, for every $i, j \leq n$, there exists $s \geq 1$ such that $M^s(i, j) > 0$. Equivalently, this means that M is not similar to a block upper-triangular matrix. We record the following general property of irreducible matrices.

LEMMA 4.6. *Let M be an irreducible matrix. Then its spectral radius ρ is a (positive) eigenvalue with a positive eingenvector. In particular, there exist positive vectors $\vec{v}_{\min}$ and $\vec{v}_{\max}$ such that, componentwise,*

$$\rho^n \vec{v}_{\min} \ \leq \ M^n \vec{1} \ \leq \ \rho^n \vec{v}_{\max} \quad \text{for all } n > 0$$

where $\vec{1}$ is the vector whose coordinates are all equal to 1. Moreover, there exist $c_{\min}, c_{\max} > 0$ such that

$$c_{\min} \rho^n \ \leq \ \vec{1}^t M^n \vec{1} \ \leq \ c_{\max} \rho^n \quad \text{for all } n > 0.$$

PROOF. We refer the reader to [**8**, chap. 13, vol. 2] for a comprehensive presentation of the properties of irreducible matrices and in particular for the Perron-Frobenius theorem, which establishes that the spectral radius of M is an eigenvalue with a positive eigenvector: let $\vec{v}_0$ be such an eigenvector, and let $\vec{v}_{\min}$ (resp. $\vec{v}_{\max}$) be appropriate multiples of $\vec{v}_0$ with all coefficients less than 1 (resp. greater than 1). Then we have, componentwise, $\rho^n \vec{v}_{\min} = M^n \vec{v}_{\min} \leq M^n \vec{1} \leq M^n \vec{v}_{\max} = \rho^n \vec{v}_{\max}$.

Let $c_{\min}$ (resp. $c_{\max}$) be the sum of the coefficients of $\vec{v}_{\min}$ (resp. $\vec{v}_{\max}$). Then, summing over all components of $M^n \vec{v}_{\min}$ and $M^n \vec{v}_{\max}$, we get $c_{\min} \rho^n \leq \vec{1}^t M^n \vec{1} \leq c_{\max} \rho^n$. □

Going back to automata, we note that a *Markov chain* can be naturally associated with a Markovian automaton: if $\mathcal{A}$ is a Markovian automaton on alphabet $\tilde{A}$, with state set Q, we define the Markov chain $M(\mathcal{A})$ on Q as follows: its transition matrix is given by $M(p, q) = \sum_{a \in \tilde{A} \text{ s.t. } p \cdot a = q} \gamma(p, a)$ for all $p, q \in Q$, and its initial vector is γ_0.

We say that the Markov chain $M(\mathcal{A})$ (or, by extension, the Markovian automaton $\mathcal{A}$), is *irreducible* if this transition matrix is irreducible, which is equivalent to the strong connectedness of $\mathcal{A}$. We note that, in that case, if $\mathcal{A}$ does not consist of a simple cycle, then $\mathcal{A}$ does not have a cycle of probability 1. In view of Proposition 4.3, this implies that the sequence of probability measures determined by $\mathcal{A}$ is prefix-heavy. We will see below (Proposition 4.9) that we can give a precise evaluation of the parameters of this sequence.

To this end, we introduce the notion of local Markovian automata, where labels can be read on states instead of edges.

More precisely a Markovian automaton is *local* if all the incoming transitions into a given state are labeled by the same letter: for all states p, q and letters a, b, if $p \cdot a = q \cdot b$ then $a = b$. If $\mathcal{A}$ is a Markovian automaton, let $\mathcal{A}'$ denote the local Markovian automaton obtained as follows.

- its set of states is $Q' = \{(q, a) \in Q \times \tilde{A} \mid \exists p \in Q,\ p \cdot a = q\}$;
- its transition function $\star$ is given by $(p, a) \star b = (q, b)$ if $p \cdot b = q$;
- its initial probability vector γ'_0 is given by

$$\gamma'_0\big((p, a)\big) = \begin{cases} \gamma_0(p) & \text{if } a \text{ is the least label of the transitions into } p \\ 0 & \text{otherwise} \end{cases}$$

  (we fix an arbitrary order on $\tilde{A}$)
- its transition probability vectors are given by $\gamma'\big((p, a), b\big) = \gamma(p, b)$.

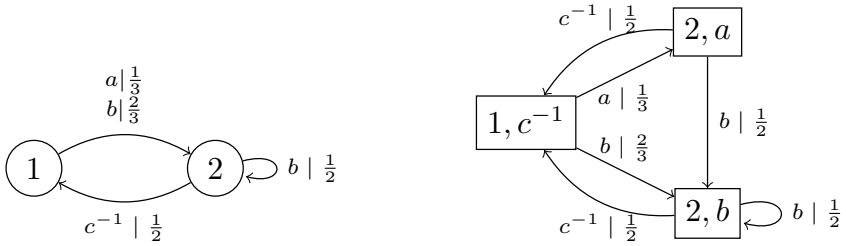

FIGURE 4. A Markovian automaton and its associated local automaton.

PROPOSITION 4.7. *Let $\mathcal{A}$ be a Markovian automaton. Then the associated local Markovian automaton $\mathcal{A}'$ assigns the same probability as $\mathcal{A}$ to every reduced word. Moreover, if $\mathcal{A}$ is irreducible, then so is $\mathcal{A}'$.*

PROOF. The first part of the statement follows directly from the definition, by a simple induction on the length of the words: indeed, we retrieve a path in $\mathcal{A}$ by forgetting the second coordinate on the states of $\mathcal{A}'$; and every path of $\mathcal{A}$ starting at some state q, can be lifted uniquely to a path in $\mathcal{A}'$ starting at any vertex of the form (q, a) of $\mathcal{A}'$.

Assume that $\mathcal{A}$ is irreducible and let (p, a) and (q, b) be states of $\mathcal{A}'$. By definition of $\mathcal{A}'$, there exists a state q' of $\mathcal{A}$ such that $q' \cdot b = q$. Moreover, since $\mathcal{A}$ is irreducible, there exists a path from p to q' in $\mathcal{A}$, say $p \xrightarrow{a_1} q_1 \xrightarrow{a_2} \dots \xrightarrow{a_t} q'$. Then

$$(p, a) \xrightarrow{a_1} (q_1, a_1) \xrightarrow{a_2} \dots \xrightarrow{a_t} (q', a_t) \xrightarrow{b} (q', b)$$

is a path in $\mathcal{A}'$ from (p, a) to (q, b), so $\mathcal{A}'$ is irreducible as well. □

If $\mathcal{A}$ is a Markovian automaton, we denote by $\mathbb{M}_{\mathcal{A}}$ (or just $\mathbb{M}$ when there is no ambiguity) the stochastic matrix associated with its local automaton $\mathcal{A}'$:

$$\mathbb{M}\big((p, a), (q, b)\big) = \begin{cases} \gamma'\big((p, a), b\big) = \gamma(p, b) & \text{if } p \cdot b = q \\ 0 & \text{otherwise.} \end{cases}$$

We also denote by $\mathbb{M}_{[2]}$ and $\mathbb{M}_{[3]}$ the matrices defined by

$$\mathbb{M}_{[2]}\big((p,a),(q,b)\big) = \Big(\mathbb{M}\big((p,a),(q,b)\big)\Big)^2 \text{ and}$$
$$\mathbb{M}_{[3]}\big((p,a),(q,b)\big) = \Big(\mathbb{M}\big((p,a),(q,b)\big)\Big)^3,$$

and by $\alpha_{[2]}$ and $\alpha_{[3]}$ the largest eigenvalue of $\mathbb{M}_{[2]}$ and $\mathbb{M}_{[3]}$, respectively. The value $\alpha_{[2]}$ is called the *coincidence probability* of $\mathcal{A}$, and it will play an important role in the sequel.

Observe that if $\mathcal{A}$ is local, then $\mathcal{A}'$ is equal to $\mathcal{A}$, up to the name of the states. We are interested in local automata for the following properties.

LEMMA 4.8. *Let $\mathcal{A}$ be a local Markovian automaton. Then the following holds*

- *for all states p, q there is at most one transition from p to q;*
- *two paths starting from the same state are labeled by the same word if and only if they go through the same states in the same order;*
- *for every $\ell \geq 0$, we have $\mathbb{M}^\ell(p,q) = \sum_{u\in\mathcal{R}_\ell, p\cdot u=q} \gamma(p,u)$, $\mathbb{M}^\ell_{[2]}(p,q) = \sum_{u\in\mathcal{R}_\ell, p\cdot u=q} \gamma(p,u)^2$ and $\mathbb{M}^\ell_{[3]}(p,q) = \sum_{u\in\mathcal{R}_\ell, p\cdot u=q} \gamma(p,u)^3$.*

We can now give an upper bound for the parameters of the sequence of probability measures determined by an irreducible Markovian automaton.

PROPOSITION 4.9. *Let $\mathcal{A}$ be an irreducible Markovian automaton with coincidence probability $\alpha_{[2]}$, and let $(\mathbb{R}_n)_n$ be the sequence of probability measures it determines. If $\mathcal{A}$ does not consist of a single cycle, then there exists a constant $C > 0$ such that $(\mathbb{R}_n)_n$ is prefix-heavy with parameters $(C, \alpha_{[2]}^{1/2})$.*

PROOF. Let v be a reduced word of length ℓ and let $q \in Q$ be a state of $\mathcal{A}$. By Lemma 4.8, we have

$$\gamma(q,v) = \sqrt{\gamma(q,v)^2} \leq \sqrt{\mathbb{M}^\ell_{[2]}(q, q\cdot v)} \leq \sqrt{\vec{1}^t \mathbb{M}^\ell_{[2]} \vec{1}}.$$

Lemma 4.6 then shows that there exists $C > 0$ such that $\gamma(q,v) \leq C\alpha_{[2]}^{\frac{\ell}{2}}$. We can now conclude as in the proof of Proposition 4.3. □

This yields the following refinement of Corollary 4.5.

COROLLARY 4.10. *Let $\mathcal{A}$ be a Markovian automaton without a probability 1 cycle and with coincidence probability $\alpha_{[2]}$. Then a tuple of reduced words of length at most n chosen at random according to $\mathcal{A}$, at $\alpha_{[2]}$-density $d < \frac{1}{8}$ (resp. $d < \frac{1}{32}$), exponentially generically has the central tree property (resp. generates a malnormal subgroup).*

4.3. Ergodic Markovian automata. If the Markovian automaton $\mathcal{A}$ is irreducible and if, in addition, for all large enough n, $M(\mathcal{A})^n(q,q) > 0$ for each $q \in Q$, we say that $\mathcal{A}$ (resp. $M(\mathcal{A})$) is *ergodic*. This is equivalent to stating that $\mathcal{A}$ has a collection of loops of relatively prime lengths, or also that all large enough integral powers of $M(\mathcal{A})$ have only positive coefficients. If $\mathcal{A}$ is ergodic, we can apply a classical theorem on Markov chains, which states that there exists a *stationary vector* $\tilde{\gamma}$ such that the distribution defined by $\mathcal{A}$ converges to that stationary vector exponentially fast (see [**18**, Thm 4.9]). In the vocabulary of Markovian automata, this yields the following theorem.

If $u \in \tilde{A}^*$ has length n, let $Q_n^p(u) = p \cdot u$ be the state of $\mathcal{A}$ reached after reading the word u starting at state p. We treat Q_n^p as a random variable.

THEOREM 4.11. *Let $\mathcal{A}$ be an ergodic Markovian automaton on alphabet $\tilde{A}$, with state set Q ($|Q| \geq 2$). For each $q \in Q$, the limit $\lim_{n\to\infty} \mathbb{R}_n[Q_n^p = q]$ exists, and if we denote it by $\tilde{\gamma}(q)$, then $\tilde{\gamma}$ is a probability vector (called the* stationary vector*). In addition, there exist $K > 0$ and $0 < c < 1$, such that $|\mathbb{R}_n[Q_n^p = q] - \tilde{\gamma}(q)| < Kc^n$ for all n large enough.*

REMARK 4.12. The constant c in Theorem 4.11 is the maximal modulus of the non-1 eigenvalues of $M(\mathcal{A})$.

EXAMPLE 4.13. The Markovian automaton discussed in Example 4.1, relative to the uniform distribution on reduced words of length n, is ergodic. Its stationary vector $\tilde{\gamma}$ is equal to γ_0 ($\tilde{\gamma}(q) = \frac{1}{2r}$ for every state q), and the constant c is $\frac{1}{2r-1}$.

On the other hand, the Markovian automaton $\mathcal{A}$ in Example 4.2 is irreducible but not ergodic (loops have even lengths), and it does not have a stationary vector.

We use Theorem 4.11 to show that, under a very mild additional hypothesis, an ergodic Markovian automaton yields a prefix-heavy sequence of measures $(\mathbb{R}_n)_n$ such that $\liminf \mathbb{R}_n(\mathcal{C}) > 0$.

PROPOSITION 4.14. *Let $\mathcal{A}$ be an ergodic Markovian automaton, with initial vector γ_0 and stationary vector $\tilde{\gamma}$ and let $(\mathbb{R}_n)_n$ be the sequence of measures it induces on reduced words. If $\sum_{a\in\tilde{A}} \gamma_0(a)\tilde{\gamma}(a^{-1}) \neq 1$, then $\liminf \mathbb{R}_n(\mathcal{C}) > 0$.*

Observe that the sum $\sum_{a\in\tilde{A}} \gamma_0(a)\tilde{\gamma}(a^{-1})$ is less than 1, since we are dealing with probability vectors, unless there exists a (necessarily single) letter a such that $\gamma_0(a) = \tilde{\gamma}(a^{-1}) = 1$.

PROOF. The set $\mathcal{C}$ of cyclically reduced words is the complement in $\mathcal{R}$ of the disjoint union of the sets $a\tilde{A}^*a^{-1}$ ($a \in \tilde{A}$). Now we have

$$\begin{aligned}
\mathbb{R}_n(a\tilde{A}^*a^{-1}) &= \sum_{p\in Q} \gamma_0(p)\gamma(p,a) \left(\sum_{|u|=n-2} \gamma(p\cdot a, u)\gamma(p\cdot(au), a^{-1}) \right) \\
&= \sum_{p\in Q} \gamma_0(p)\gamma(p,a) \left(\sum_{q\in Q} \mathbb{R}_n(Q_{n-2}^{p\cdot a} = q)\gamma(q, a^{-1}) \right) \\
&= \sum_{p\in Q} \gamma_0(p)\gamma(p,a) \left(\sum_{q\in Q} (\tilde{\gamma}(q) + \varepsilon(q,n))\gamma(q, a^{-1}) \right),
\end{aligned}$$

where $|\varepsilon(q,n)| \leq Kc^{n-2}$, with K and c given by Theorem 4.11. Then we have

$$\mathbb{R}_n(a\tilde{A}^*a^{-1}) = \gamma_0(a)\tilde{\gamma}(a^{-1}) + \gamma_0(a) \left(\sum_{q\in Q} \varepsilon(q,n)\gamma(q,a^{-1}) \right)$$

and $\lim \mathbb{R}_n(a\tilde{A}^*a^{-1}) = \gamma_0(a)\tilde{\gamma}(a^{-1})$. It follows that

$$\lim \mathbb{R}_n(\mathcal{C}) = 1 - \sum_{a\in\tilde{A}} \gamma_0(a)\tilde{\gamma}(a^{-1}),$$

thus concluding the proof. □

Proceeding as in Section 3.6, we can use Proposition 4.14, Corollary 3.14 and the results of Section 3.5, to generalize part of Theorem 2.4 (2), and show that, up to $\alpha_{[2]}$-density $\frac{\lambda}{4}$, a tuple of cyclically reduced words of length at most n chosen at random according to $\mathcal{A}$, exponentially generically satisfies the small cancellation property $C'(\lambda)$. We will now see (Theorem 4.15) that we can improve this bound, and go up to $\alpha_{[2]}$-density $\frac{\lambda}{2}$.

4.4. Phase transitions for the Markovian model. We can now state a phase transition theorem, which generalizes parts of Theorem 2.4. Let us say that an ergodic Markovian automaton is *non-degenerate* if its initial distribution γ_0 and its stationary vector $\tilde{\gamma}$ satisfy $\sum_{a\in\tilde{A}} \gamma_0(a)\tilde{\gamma}(a^{-1}) \neq 1$.

THEOREM 4.15. *Let $\mathcal{A}$ be a non-degenerate ergodic Markovian automaton with coincidence probability $\alpha_{[2]}$. Let $0 < d < 1$ and let G be the group presented by a tuple $\vec{h}$ of cyclically reduced words of length n, chosen independently and at random according to $\mathcal{A}$, at $\alpha_{[2]}$-density d. Then we have the following phase transitions:*

- *if $0 < \lambda < \frac{1}{2}$ and $0 < d < \frac{\lambda}{2}$, then exponentially generically $\vec{h}$ satisfies the small cancellation property $C'(\lambda)$; if $\lambda = \frac{1}{6}$, then G is generically infinite and hyperbolic;*
- *if $d > \frac{\lambda}{2}$ then exponentially generically $\vec{h}$ does not satisfy the small cancellation property $C'(\lambda)$;*
- *if $d > \frac{1}{2}$ then exponentially generically G is degenerated in a sense that is made precise in Proposition 4.23, and which implies that G is a free group or the free product of a free group with $\mathbb{Z}/2\mathbb{Z}$.*

The rest of the paper is devoted to the proof of Theorem 4.15. The first statement is established in Proposition 4.16, while the second and third statements are proved respectively in Propositions 4.22 and 4.23.

4.5. Long common factors at low density. In this section we estimate the probability that random words share a long common factor. More precisely, we show the following statement, the first part of Theorem 4.15.

PROPOSITION 4.16. *Let $\mathcal{A}$ be a non-degenerate ergodic Markovian automaton with coincidence probability $\alpha_{[2]}$. Let $\lambda \in (0, \frac{1}{2})$ and let $d \in (0, \frac{\lambda}{2})$. A tuple of cyclically reduced words of length n taken independently and randomly according to $\mathcal{A}$, at $\alpha_{[2]}$-density d, exponentially generically satisfies the small cancellation property $C'(\lambda)$.*

The structure of the proof of Proposition 4.16 resembles that of the proof of Theorem 3.20, and requires the consideration of several cases. This is the object of the rest of Section 4.5.

To this end, we introduce additional notation: let $\vec{\gamma}_q(n)$ be the vector of coordinates $\gamma(q,u)$ when u ranges over $\mathcal{R}_n$ in lexicographic order, and let $\|\vec{\gamma}_q(n)\|_k = (\sum_{u\in\mathcal{R}_n} \gamma(q,u)^k)^{1/k}$ be the ℓ_k-norm of this vector. We start with an elementary result.

LEMMA 4.17. *Let $\mathcal{A}$ be a Markovian automaton, let $0 < i, \ell < n$ be integers, and let $u \in \mathcal{R}_\ell$. The probability $\mathfrak{p}$ that u occurs as a cyclic factor at position i in a reduced word of length n is bounded above by*

$$\begin{cases} \sum_{q\in Q} \gamma(q,u) & \text{if } i \leq n-\ell+1 \\ \sum_{q,q'\in Q} \gamma(q,u_1)\gamma(q',u_2) & \text{if } i > n-\ell+1 \text{ and } u = u_1u_2 \text{ with } |u_1| = n-i+1 \end{cases}$$

PROOF. If $i \leq n-\ell+1$, then $\mathfrak{p} = \mathbb{R}_n(\tilde{A}^{i-1}u\tilde{A}^{n-\ell-i+1})$ is equal to

$$\begin{aligned}\sum_{p\in Q}\gamma_0(p)\sum_{w\in\mathcal{R}_{i-1}}\gamma(p,w)\gamma(p\cdot w,u) &= \sum_{p\in Q}\gamma_0(p)\sum_{q\in Q}\sum_{\substack{w\in\mathcal{R}_{i-1}\\ p\cdot w=q}}\gamma(p,w)\gamma(q,u)\\ &= \sum_{p\in Q}\gamma_0(p)\sum_{q\in Q}\mathbb{R}_{i-1}[Q_{i-1}^p=q]\gamma(q,u)\\ &\leq \sum_{p,q\in Q}\gamma_0(p)\gamma(q,u) = \sum_{q\in Q}\gamma(q,u).\end{aligned}$$

If $i > n-\ell+1$ and $u = u_1u_2$ with $|u_1| = n-i+1$, then

$$\begin{aligned}\mathfrak{p} = \mathbb{R}_n(u_2\tilde{A}^{n-\ell}u_1) &= \sum_{q'\in Q}\gamma_0(q')\gamma(q',u_2)\sum_{w\in\mathcal{R}_{n-\ell}}\gamma(q'\cdot u_2,w)\gamma(q'\cdot u_2w,u_1)\\ &= \sum_{q'\in Q}\gamma_0(q')\gamma(q',u_2)\sum_{q\in Q}\sum_{\substack{w\in\mathcal{R}_{n-\ell}\\ q'\cdot u_2w=q}}\gamma(q'\cdot u_2,w)\gamma(q,u_1)\\ &= \sum_{q'\in Q}\gamma_0(q')\gamma(q',u_2)\sum_{q\in Q}\mathbb{R}_{n-\ell}[Q_{n-\ell}^{q'\cdot u_2}=q]\gamma(q,u_1)\\ &\leq \sum_{q,q'\in Q}\gamma(q,u_1)\gamma(q',u_2),\end{aligned}$$

which concludes the proof. □

PROPOSITION 4.18. *Let $\mathcal{A}$ be an irreducible Markovian automaton with coincidence probability $\alpha_{[2]}$. Let n, ℓ, i and j be positive integers such that $\ell \leq n$ and $i,j \leq n$. Denote by $L(n,\ell,i,j)$ the probability that two reduced words of length n share a common cyclic factor of length ℓ at positions respectively i and j. Then there exists a positive constant K such that*

$$L(n,\ell,i,j) \leq K\alpha_{[2]}^{\ell}.$$

PROOF. Without loss of generality (see Proposition 4.7), we may assume that $\mathcal{A}$ is local. The proof is based on a case study.
Case 1: $i,j \leq n-\ell+1$. Using Lemma 4.17, we have

$$L(n,\ell,i,j) \leq \sum_{p,q\in Q}\sum_{u\in\mathcal{R}_\ell}\gamma(p,u)\gamma(q,u).$$

By a repeated application of the Cauchy-Schwarz inequality, we get

$$L(n,\ell,i,j) \;\leq\; \sum_{p,q\in Q}\|\vec{\gamma}_p(\ell)\|_2\|\vec{\gamma}_q(\ell)\|_2 \;\leq\; \sum_{q\in Q}\|\vec{\gamma}_q(\ell)\|_2^2. \tag{5}$$

Now, in view of Lemma 4.8 and since $\mathcal{A}$ is local, we have

$$\sum_{q\in Q}\|\vec{\gamma}_q(\ell)\|_2^2 = \sum_{q\in Q}\sum_{u\in\mathcal{R}_\ell}\gamma(q,u)^2 = \sum_{p\in Q}\sum_{q\in Q}\sum_{\substack{u\in\mathcal{R}_\ell\\ p\cdot u=q}}\gamma(p,u)^2 = \vec{1}^t\mathbb{M}_{[2]}^{\ell}\vec{1}. \tag{6}$$

Since $\mathbb{M}$ is irreducible, Lemma 4.6 shows that there exists a positive constant $K>0$ such that, for ℓ large enough, we have

$$L(n,\ell,i,j) \;\leq\; \sum_{q\in Q}\|\vec{\gamma}_q(\ell)\|_2^2 \;=\; \vec{1}^t\mathbb{M}_{[2]}^{\ell}\vec{1} \;\leq\; K\alpha_{[2]}^{\ell},$$

which concludes the proof of the statement in that case.

Case 2: $i > n-\ell+1$ and $j \leq n-\ell+1$. (The case where $i \leq n-\ell+1$ and $j > n-\ell+1$ is symmetrical.) Let $k = n-i+1$ (so $1 \leq k < \ell$). By Lemma 4.17, we have

$$\begin{aligned} L(n,\ell,i,j) &\leq \sum_{\substack{u_1\in\mathcal{R}_k\\ u_2\in\mathcal{R}_{\ell-k}}} \sum_{p,p',q\in Q} \gamma(p,u_1)\gamma(p',u_2)\gamma(q,u_1u_2) \\ &\leq \sum_{\substack{u_1\in\mathcal{R}_k\\ u_2\in\mathcal{R}_{\ell-k}}} \sum_{p,p',q,q'\in Q} \gamma(p,u_1)\gamma(p',u_2)\gamma(q,u_1)\gamma(q',u_2) \\ &\leq \left(\sum_{u_1\in\mathcal{R}_k}\sum_{p,q\in Q}\gamma(p,u_1)\gamma(q,u_1)\right)\left(\sum_{u_2\in\mathcal{R}_{\ell-k}}\sum_{p',q'\in Q}\gamma(p',u_2)\gamma(q',u_2)\right). \end{aligned}$$

By Cauchy-Schwarz, it follows that

$$\begin{aligned} L(n,\ell,i,j) &\leq \left(\sum_{p,q\in Q}\|\vec{\gamma}_p(k)\|_2\|\vec{\gamma}_q(k)\|_2\right)\left(\sum_{p',q'\in Q}\|\vec{\gamma}_{p'}(\ell-k)\|_2\|\vec{\gamma}_{q'}(\ell-k)\|_2\right) \\ &\leq \left(\sum_{q\in Q}\|\vec{\gamma}_q(k)\|_2^2\right)\left(\sum_{q\in Q}\|\vec{\gamma}_q(\ell-k)\|_2^2\right) \\ &\leq \left(\vec{1}^t\ \mathbb{M}_{[2]}^k\ \vec{1}\right)\left(\vec{1}^t\ \mathbb{M}_{[2]}^{\ell-k}\ \vec{1}\right) \text{ by Equation (6).} \end{aligned}$$

By Lemma 4.6, there exists a constant K_1 such that these two factors are bounded above, respectively, by $K_1\alpha_{[2]}^k$ and $K_1\alpha_{[2]}^{\ell-k}$. Therefore

$$L(n,\ell,i,j) \leq K_1^2\alpha_{[2]}^\ell$$

as announced.

Case 3: $i,j > n-\ell+1$. Without loss of generality, we may assume that $i<j$, and we let $k = n-j+1$ and $k' = \ell-(n-i+1)$. Then a word u of length ℓ occurs as a cyclic factor in two reduced words w_1 and w_2 of length n, at positions i and j respectively, if $u = u_1u_2u_3$ with $|u_1| = k$, $|u_2| = j-i$ and $|u_3| = k'$, and if $w_1 \in u_3\tilde{A}^{n-\ell}u_1u_2$ and $w_2 \in u_2u_3\tilde{A}^{n-\ell}u_1$. Then we have

$$\begin{aligned} L(n,\ell,i,j) &\leq \sum_{\substack{u_1\in\mathcal{R}_k\\ u_2\in\mathcal{R}_{j-i}\\ u_3\in\mathcal{R}_{k'}}} \sum_{\substack{p,p'\in Q\\ q,q''\in Q}} \gamma(q,u_1u_2)\gamma(q'',u_3)\ \gamma(p,u_1)\gamma(p',u_2u_3) \\ &\leq \sum_{\substack{u_1\in\mathcal{R}_k\\ u_2\in\mathcal{R}_{j-i}\\ u_3\in\mathcal{R}_{k'}}} \sum_{\substack{p,p',p''\in Q\\ q,q',q''\in Q}} \gamma(q,u_1)\gamma(q',u_2)\gamma(q'',u_3)\ \gamma(p,u_1)\gamma(p',u_2)\gamma(p'',u_3) \\ &\leq \sum_{\substack{u_1\in\mathcal{R}_k\\ p',q\in Q}} \gamma(q,u_1)\gamma(p',u_1) \sum_{\substack{u_2\in\mathcal{R}_{j-i}\\ p,q''\in Q}} \gamma(q'',u_2)\gamma(p,u_2) \sum_{\substack{u_3\in\mathcal{R}_{k'}\\ p'',q'\in Q}} \gamma(q',u_3)\gamma(p'',u_3). \end{aligned}$$

By the Cauchy-Schwarz inequality, $L(n,\ell,i,j)$ is at most equal to

$$\sum_{p,q\in Q}\|\vec{\gamma}_p(k)\|_2\|\vec{\gamma}_q(k)\|_2 \sum_{p,q\in Q}\|\vec{\gamma}_p(j-i)\|_2\|\vec{\gamma}_q(j-i)\|_2 \sum_{p,q\in Q}\|\vec{\gamma}_p(k')\|_2\|\vec{\gamma}_q(k')\|_2$$

and hence to

$$\sum_{q\in Q}\|\vec{\gamma}_q(k)\|_2^2\ \sum_{q\in Q}\|\vec{\gamma}_q(j-i)\|_2^2\ \sum_{q\in Q}\|\vec{\gamma}_q(k')\|_2^2.$$

Lemma 4.6 shows that these three factors are bounded above, respectively, by $K_1\alpha_{[2]}^k$, $K_1\alpha_{[2]}^{j-i}$ and $K_1\alpha_{[2]}^{k'}$ for some constant K_1. Therefore

$$L(n,\ell,i,j)\le K_1^3\alpha_{[2]}^{k+j-i+k'}=K_1^3\alpha_{[2]}^{\ell},$$

as announced. □

PROPOSITION 4.19. *Let $\mathcal{A}$ be an irreducible Markovian automaton with coincidence probability $\alpha_{[2]}$. Denote by $L^{(2)}(n,\ell,i,j)$ the probability for two reduced words of length n to have an occurrence of a factor of length ℓ in the first word at position i, and an occurrence of its inverse in the second word, at position j, with $\ell\le n$ and $i,j\le n-\ell+1$. Then there exists a positive constant K such that*

$$L^{(2)}(n,\ell,i,j)\le K\alpha_{[2]}^{\ell}.$$

PROOF. The proof follows the same steps as that of Proposition 4.18. In the first case ($i,j\le n-\ell+1$), Lemma 4.17 shows that

$$L^{(2)}(n,\ell,i,j)\le\sum_{p,q\in Q}\sum_{u\in\mathcal{R}_\ell}\gamma(p,u)\gamma(q,u^{-1}).$$

Since the set of reduced words of length ℓ and the set of their inverses are equal, we get, by the Cauchy-Schwarz inequality,

$$L^{(2)}(n,\ell,i,j)\le\sum_{p,q\in Q}\|\vec{\gamma}_p(\ell)\|_2\|\vec{\gamma}_q(\ell)\|_2,$$

and the proof proceeds as in the corresponding case of Lemma 4.18.

In the second case ($i>n-\ell+1$ and $j\le n-\ell+1$), if $k=n-i+1$, then we have

$$\begin{aligned}L^{(2)}(n,\ell,i,j)&\le\sum_{\substack{u_1\in\mathcal{R}_k\\u_2\in\mathcal{R}_{\ell-k}}}\sum_{p,p',q\in Q}\gamma(p,u_1)\gamma(p',u_2)\gamma(q,u_2^{-1}u_1^{-1})\\&\le\sum_{\substack{u_1\in\mathcal{R}_k\\u_2\in\mathcal{R}_{\ell-k}}}\sum_{p,p',q,q'\in Q}\gamma(p,u_1)\gamma(p',u_2)\gamma(q,u_2^{-1})\gamma(q',u_1^{-1})\\&\le\Big(\sum_{u_1\in\mathcal{R}_k}\sum_{p,q'\in Q}\gamma(p,u_1)\gamma(q',u_1^{-1})\Big)\Big(\sum_{u_2\in\mathcal{R}_{\ell-k}}\sum_{p',q\in Q}\gamma(p',u_2)\gamma(q,u_2^{-1})\Big)\end{aligned}$$

and as in the previous case, the proof proceeds as in Lemma 4.18.

The situation is a little more complex in the last case ($i,j>n-\ell+1$). Without loss of generality, we may assume that $i<j$. With the same notation as in the proof of Lemma 4.18, we distinguish two cases. If $|u_3|<|u_2|$ (that is, $\ell-k<k'$, or $\ell+i+j<2n+2$), we let $u_2=u_2'u_2''$ with $|u_2'|=|u_3|$. Then $w_1\in u_3\tilde{A}^{n-\ell}u_1u'2u_2''$ and $w_2\in {u_2'}^{-1}u_1^{-1}\tilde{A}^{n-\ell}u_3^{-1}{u_2''}^{-1}$ and, as in the previous proof, we find that $L^{(2)}(n,\ell,i,j)$ is at most equal to the sum of the

$$\gamma(p,u_1)\gamma(q,u_1^{-1})\gamma(p',u_2')\gamma(q',{u_2'}^{-1})\gamma(p'',u_2'')\gamma(q'',{u_2''}^{-1})\gamma(p''',u_3)\gamma(q''',u_3^{-1})$$

with $u_1\in\mathcal{R}_{j-i}$, $u_2'\in\mathcal{R}_{\ell-k}$, $u_2''\in\mathcal{R}_{k'-(\ell-k)}$, $u_3\in\mathcal{R}_{\ell-k}$, and $p,p',p'',p''',q,q',q'',q'''$ are states in Q. The proof then proceeds as before, with multiple applications of the Cauchy-Schwarz inequality.

The case where $|u_3| \geq |u_2|$ (that is, $\ell + i + j \geq 2n + 2$) is handled in the same fashion. □

COROLLARY 4.20. *Let $\mathcal{A}$ be a non-degenerated ergodic Markovian automaton with coincidence probability $\alpha_{[2]}$. Let n, ℓ, i, j be positive integers such that $\ell \leq n$ and $i, j \leq n$. There exists a constant $K > 0$ such that the probability $\mathfrak{p}$ that two cyclically reduced words of length n have occurrences of the same word of length ℓ (resp. of a word of length ℓ and its inverse) as cyclic factors at positions respectively i and j, satisfies $\mathfrak{p} \leq K\alpha_{[2]}^{\ell}$.*

PROOF. The hypothesis on $\mathcal{A}$ guarantees that $\liminf \mathbb{R}_n(\mathcal{C}) = p > 0$ by Proposition 4.14. Our statement then follows from Propositions 4.18 and 4.19, in view of Lemma 3.13. □

We now consider the case of multiple occurrences of a length ℓ cyclic factor (or of such a word and its inverse) within a single reduced word.

PROPOSITION 4.21. *Let $\mathcal{A}$ be a non-degenerate ergodic Markovian automaton with coincidence probability $\alpha_{[2]}$. There exists a constant $K > 0$ such that the probability that a cyclically reduced word of length n has two occurrences of a length ℓ word as cyclic factors, or occurrences of a length ℓ word and its inverse as cyclic vactors, is at most $K\ell^2 n^2 \alpha_{[2]}^{\ell/2}$.*

PROOF. By Proposition 4.9, the sequence $(\mathbb{R}_n)_n$ induced by $\mathcal{A}$ is prefix-heavy with parameters $(C, \alpha_{[2]}^{1/2})$ for some C. The result then follows from Corollary 3.14. □

We can now proceed with the **proof of Proposition 4.16.** Let $N = \alpha_{[2]}^{-dn}$. An N-tuple of cyclically reduced words which fails to satisfy $C'(\lambda)$, must satisfy one of the following conditions: either two words in the tuple have occurrences of the same cyclic factor of length $\ell = \lambda n$ or occurrences of such a word and its inverse; or a word in the tuple has two occurrences of the same cyclic factor of length ℓ or occurrences of such a word and its inverse.

By Corollary 4.20, the first event occurs with probability at most

$$K\binom{N}{2} n^2 \alpha_{[2]}^{\ell} \leq K n^2 \alpha_{[2]}^{(\lambda - 2d)n}$$

for some $K > 0$. By Proposition 4.21, the second event occurs with probability at most

$$KN\ell^2 n^2 \alpha_{[2]}^{\frac{\ell}{2}} \leq K n^4 \alpha_{[2]}^{(\frac{\lambda}{2} - d)n},$$

for some $K > 0$. Thus both events occur with probabilities that vanish exponentially fast, and this concludes the proof of Proposition 4.16.

4.6. Long common prefixes at high density. In this section, we establish the following propositions corresponding respectively to the second and third statement of Theorem 4.15.

PROPOSITION 4.22. *Let $\mathcal{A}$ be a non-degenerate ergodic Markovian automaton with coincidence probability $\alpha_{[2]}$. Let $\lambda \in (0, \frac{1}{2})$ and let $d \in (\frac{\lambda}{2}, 1)$. A tuple of cyclically reduced words of length n taken independently and randomly according to $\mathcal{A}$, at density d, generically does not satisfy the small cancellation property $C'(\lambda)$.*

PROPOSITION 4.23. *Let $\mathcal{A}$ be a non-degenerate ergodic Markovian automaton with coincidence probability $\alpha_{[2]}$. Let E be the set of letters of $\tilde{A}$ which label a transition in $\mathcal{A}$ and let $D = A \setminus (E \cup E^{-1})$. Let $d > \frac{1}{2}$ and $N \geq \alpha_{[2]}^{-dn}$, and let G be a group presented by an N-tuple of cyclically reduced words chosen independently at random according to $\mathcal{A}$.*

If $E \cap E^{-1} = \emptyset$, then $G = F(|D| + 1)$ exponentially generically.

If $E \cap E^{-1} \neq \emptyset$, then exponentially generically $G = F(D) * \mathbb{Z}/2\mathbb{Z}$ (if n is even) or $G = F(D)$ (if n is odd).

Both proofs rely heavily on the methodology introduced by Szpankowski [**32**] to study the typical heigth of a random trie. We first establish simple lower and upper bounds for words to share a common prefix (Lemmas 4.24 and 4.25).

LEMMA 4.24. *Let $\mathcal{A}$ be an irreducible Markovian automaton with coincidence probability $\alpha_{[2]}$. Let $P(n, \ell)$ (*resp. *$P'(n, \ell)$) be the probability that two reduced (resp. cyclically reduced) words of length n share a common prefix of length ℓ. There exists a constant $K > 0$ such that $P(n, \ell) \geq K\alpha_{[2]}^{\ell}$.*

If $\mathcal{A}$ is non-degenerate and ergodic and t is large enough for all the coefficients of $\mathbb{M}^t$ to be positive, then K can be chosen such that $P'(n, \ell) \geq K\alpha_{[2]}^{\ell}$ when $n \geq \ell + t + 1$.

PROOF. Let p be a state such that $\gamma_0(p) > 0$. To establish the announced lower bounds, we only need to consider the words that can be read from state p. More precisely, when considering reduced words, we have

$$P(n, \ell) \;\geq\; \gamma_0(p)^2 \sum_{u \in \mathcal{R}_\ell} \gamma(p, u)^2.$$

We observe that $\sum_{u \in \mathcal{R}_\ell} \gamma(p, u)^2$ is the p-component of $\mathbb{M}_{[2]}^{\ell}\vec{1}$, and by Lemma 4.6, it is greater than or equal to $\beta\alpha_{[2]}^{\ell}$, where β is the minimal component of $\vec{v}_{\min}$ (in the notation of Lemma 4.6). This completes the proof of the statement concerning $P(n, \ell)$.

We now consider cyclically reduced words, under the hypothesis that $\mathcal{A}$ is non-degenerate and ergodic. Let t be such that all the coefficients of $\mathbb{M}^t$ are positive, let $\bar{p}_{\min}$ be the least coefficient of this matrix, and let $p_{\min}$ be the least positive coefficient of $\mathbb{M}$. Finally, let $\mathfrak{p} = \liminf \mathbb{R}_n(\mathcal{C})$, which is positive by Proposition 4.14. Let X (resp. X_p) be the set of pairs of cyclically reduced words of length n that have a common prefix of length ℓ (resp. which can be read from state p). We note that

$$P'(n, \ell) \;=\; \frac{\mathbb{R}_n(X)}{\mathbb{R}_n(\mathcal{C})^2} \;\geq\; \frac{1}{\mathfrak{p}^2}\mathbb{R}_n(X) \;\geq\; \frac{1}{\mathfrak{p}^2}\mathbb{R}_n(X_p),$$

so we only need to find a lower bound for $\mathbb{R}_n(X_p)$.

Suppose that $n \geq \ell + t + 1$. Then X_p contains the set of pairs of reduced words of the form $(uu_1u_1'a, uu_2u_2'a)$ which can be read from p, where a is the first letter of u, and u_1' and u_2' are words of length t such that $p \cdot (uu_1u_1') = p \cdot (uu_2u_2') = p$. Since these words start and end with the same letters, they are guaranteed to be cyclically reduced. Thus we have

$$\mathbb{R}_n(X_p) \;\geq\; \gamma_0(p)^2 \sum_{u \in \mathcal{R}_\ell} \gamma(p, u)^2 \; p_{\min}^2 \; \bar{p}_{\min}^2 \;\geq\; \beta \; \gamma_0(p)^2 \; p_{\min}^2 \; \bar{p}_{\min}^2 \; \alpha_{[2]}^{\ell},$$

and this concludes the proof. □

LEMMA 4.25. *Let $\mathcal{A}$ be an irreducible Markovian automaton with coincidence probability $\alpha_{[2]}$. There exists a constant $K > 0$ such that the probability that three reduced words share the same prefix of length ℓ is at most $K\alpha_{[3]}^{\ell}$.*

If $\mathcal{A}$ is non-degenerate and ergodic, the same holds for triples of cyclically reduced words.

PROOF. The probability $\mathfrak{p}(u)$ that three reduced words have a common prefix u is

$$\mathfrak{p}(u) = \sum_{p_1,p_2,p_3 \in Q} \gamma_0(p_1)\ \gamma_0(p_2)\ \gamma_0(p_3)\ \gamma(p_1,u)\ \gamma(p_2,u)\ \gamma(p_3,u).$$

The probability we are interested in is obtained by summing over all $u \in \mathcal{R}_\ell$. It is bounded above by

$$\sum_{p_1,p_2,p_3 \in Q} \sum_{u \in \mathcal{R}_\ell} \gamma(p_1,u)\ \gamma(p_2,u)\ \gamma(p_3,u).$$

By the Hölder and Cauchy-Schwarz inequalities, we have

$$\begin{aligned}
&\sum_{u \in \mathcal{R}_\ell} \gamma(p_1,u)\ \gamma(p_2,u)\ \gamma(p_3,u) \\
&\qquad \le \left(\sum_{u \in \mathcal{R}_\ell} \gamma(p_1,u)^3 \right)^{\frac{1}{3}} \left(\sum_{u \in \mathcal{R}_\ell} \gamma(p_2,u)^{\frac{3}{2}}\ \gamma(p_3,u)^{\frac{3}{2}} \right)^{\frac{2}{3}} \\
&\qquad \le \left(\sum_{u \in \mathcal{R}_\ell} \gamma(p_1,u)^3 \right)^{\frac{1}{3}} \left(\sum_{u \in \mathcal{R}_\ell} \gamma(p_2,u)^3 \right)^{\frac{1}{3}} \left(\sum_{u \in \mathcal{R}_\ell} \gamma(p_3,u)^3 \right)^{\frac{1}{3}}.
\end{aligned}$$

Moreover, we have

$$\sum_{p \in Q} \sum_{u \in \mathcal{R}_\ell} \gamma(p_1,u)^3 = \vec{1}^t\ \mathbb{M}_{[3]}^{\ell}\ \vec{1}.$$

We now get the announced result using Lemma 4.6, Lemma 4.8 and the spectral properties of $\mathbb{M}_{[3]}^{\ell}$. The generalisation to cyclically reduced words follows from Lemma 3.13. □

We now build on the previous lemmas to show that, exponentially generically, large tuples of cyclically reduced words contain pairs of words with a common prefix of a prescribed length.

PROPOSITION 4.26. *Let $\mathcal{A}$ be an irreducible Markovian automaton with coincidence probability $\alpha_{[2]}$. Let $(\ell_n)_n$ be an unbounded, monotonous sequence of positive integers such that $\ell_n \le n$ for each n, and let $d > \frac{1}{2}$. Then an $\alpha_{[2]}^{-d\ell_n}$-tuple of reduced words of length n drawn randomly according to $\mathcal{A}$ generically contains two words with the same prefix of length ℓ_n.*

If $\mathcal{A}$ is non-degenerate and ergodic, the same holds for $\alpha_{[2]}^{-d\ell_n}$-tuples of cyclically reduced words.

PROOF. We use the so-called *second moment method*, as developed in [**32**], and we introduce the following notation to this end. Since the results of [**32**] are established for right-infinite words, we need to considered such words first; the result on words of length n directly follows by truncation. A right-infinite reduced word is an element u of $\tilde{A}^{\mathbb{N}}$ such that for every $i \in \mathbb{N}$, $u_i \neq u_{i+1}^{-1}$. We define the probability distribution $\mathbb{R}_\infty$ on right-infinite words induced by the Markovian automaton $\mathcal{A}$ by

first setting $\mathbb{R}_\infty(\mathcal{P}_\infty(u)) = \gamma(u)$, where $\mathcal{P}_\infty(u)$ is the set of right-infinite reduced words w such that the finite reduced word u is a prefix of w. The probability is then extended to the σ-algebra generated by the $\mathcal{P}_\infty(u)$, when u ranges over all finite reduced words (see [**34**] for more details on this kind of constructions). Let $N = \alpha_{[2]}^{-d\ell_n}$ and consider an N-tuple $\vec{h} = (h_i)_{1\leq i\leq N}$ of right-infinite reduced words, independently and randomly generated according to $\mathcal{A}$.

For $1 \leq i < j \leq N$, let $X_{i,j}$ be the random variable computing the length of the longest common prefix of h_i and h_j. We want to show that, exponentially generically,

$$\max_{1\leq i<j\leq N} X_{i,j} \geq \ell_n.$$

Let us relabel the random variables $X_{i,j}$ $(i \neq j)$ as $Y_1, \ldots, Y_m$, with $m = \binom{N}{2}$ and, say, $Y_1 = X_{1,2}$. We are therefore computing the maximum of m random variables, which are identically distributed but not independent. Fortunately, they behave almost as if they were independent, as we will see.

Let d' be such that $\frac{1}{2} < d' < d$ and for each $m \geq 1$, let

$$r_m = \log_{\alpha_{[2]}^{-2d'}}(m) = \frac{\log\binom{N}{2}}{\log \alpha_{[2]}^{-2d'}} \sim \frac{\log \alpha_{[2]}^{-2d\ell_n}}{\log \alpha_{[2]}^{-2d'}} = \frac{d\ell_n}{d'}.$$

In particular, r_m is asymptotically greater than ℓ_n, and we only need to show that

$$\lim_{n\to\infty} \mathbb{R}_\infty\left(\max_{k\in[m]} Y_k \geq r_m\right) = 1. \tag{7}$$

Let $\nu(r_m)$ denote the quantity

$$\nu(r_m) = \sum_{k=2}^{m} \frac{\mathbb{R}_\infty(Y_1 \geq r_m, Y_k \geq r_m)}{m\,\mathbb{R}_\infty(Y_1 \geq r_m)^2}.$$

We use Lemma 3 in [**32**], which states that the desired equation (7) holds if

$$\lim_{n\to\infty} m\,\mathbb{R}_\infty(Y_1 > r_m) = +\infty \text{ and } \lim_{n\to\infty} \nu(r_m) = 1.$$

We now proceed with the proof of these two equalities. By Lemma 4.24, we have $\mathbb{R}_\infty(Y_1 \geq r_m) \geq K\,\alpha_{[2]}^{r_m}$. Then

$$\begin{aligned}\log\left(m\mathbb{R}_\infty(Y_1 \geq r_m)\right) &\geq \log m + \log K + r_m \log \alpha_{[2]} \\ &= r_m \log(\alpha_{[2]}^{-2d'}) + \log K + r_m \log \alpha_{[2]} \\ &= r_m \log(\alpha_{[2]}^{1-2d'}) + \log K,\end{aligned}$$

which tends to $+\infty$, since $1 - 2d' < 0$ and $\alpha_{[2]} < 1$. Therefore,

$$\lim_{n\to\infty} m\,\mathbb{R}_\infty(Y_1 \geq r_m) = +\infty.$$

Let us now consider $\nu(r_m)$. Note that, if the Y_i were independent random variables, we would have $\nu(r_m) = \frac{m-1}{m}$, which tends to 1 when n tends to ∞.

Observe that if $2 < i < j \leq N$, then $X_{1,2}$ and $X_{i,j}$ are independant and identically distributed, so

$$\mathbb{R}_\infty(X_{1,2} \geq r_m, X_{i,j} \geq r_m) = \mathbb{R}_\infty(X_{1,2} \geq r_m)\,\mathbb{R}_\infty(X_{i,j} \geq r_m) = \mathbb{R}_\infty(Y_1 \geq r_m)^2.$$

Also, since h_1 and h_2 are drawn independently, we have $\mathbb{R}_\infty(X_{1,2} \geq r_m, X_{1,k} \geq r_m) = \mathbb{R}_\infty(X_{1,2} \geq r_m, X_{2,k} \geq r_m)$ for each $k \geq 3$. Therefore

$$\nu(r_m) = 2 \sum_{k=3}^{N} \frac{\mathbb{R}_\infty(X_{1,2} \geq r_m, X_{1,k} \geq r_m)}{m\, \mathbb{R}_\infty(Y_1 \geq r_m)^2} + \binom{N-2}{2} \frac{1}{m}.$$

Since $m = \binom{N}{2}$, we have $\lim_n \binom{N-2}{2} \frac{1}{m} = 1$. Moreover, the joint probability $\mathbb{R}_\infty(X_{1,2} \geq r_m, X_{1,k} \geq r_m)$ is exactly the probability that three random reduced words share a common prefix of length r_m: by Lemma 4.25, this is at most equal to $K\, \alpha_{[3]}^{r_m}$ for some constant $K > 0$. Together with Lemma 4.24, this yields

$$\sum_{k=3}^{N} \frac{\mathbb{R}_\infty(X_{1,2} \geq r_m, X_{1,k} \geq r_m)}{m\, \mathbb{R}_\infty(Y_1 \geq r_m)^2} \leq \frac{K'}{N} \left(\frac{\alpha_{[3]}}{\alpha_{[2]}^2} \right)^{r_m},$$

for some $K' > 0$. In [**16**] it is proved that $(\alpha_{[m]})^{1/m}$ is a decreasing sequence, so we have $\alpha_{[3]}^{1/3} \leq \alpha_{[2]}^{1/2}$ and hence

$$\left(\frac{\alpha_{[3]}}{\alpha_{[2]}^2} \right)^{r_m} \leq \left(\frac{\alpha_{[2]}^{3/2}}{\alpha_{[2]}^2} \right)^{r_m} \leq \alpha_{[2]}^{-\frac{r_m}{2}}.$$

Therefore

$$\log \left(\frac{1}{N} \left(\frac{\alpha_{[3]}}{\alpha_{[2]}^2} \right)^{r_m} \right) = -\log N - \frac{r_m}{2} \log \alpha_{[2]} \leq -\frac{1}{2} \log m + K'' - \frac{r_m}{2} \log \alpha_{[2]}$$

for some constant K''. By definition of r_m, we have $\log m = -2d' r_m \log \alpha_{[2]}$ and it follows that

$$\log \left(\frac{1}{N} \left(\frac{\alpha_{[3]}}{\alpha_{[2]}^2} \right)^{r_m} \right) \leq \frac{r_m}{2} (2d' - 1) \log \alpha_{[2]} + K''.$$

This quantity tends to $-\infty$ when n tends to ∞ since $2d' - 1 > 0$ and $\alpha_{[2]} < 1$. This proves finally that $\lim_{m \to \infty} \nu(r_m) = 1$ and establishes Equation (7). That is, the desired statement is proved for tuples of infinite reduced words. As $\ell_n \leq n$, considering right-infinite words and truncating then at their prefix of length n yields the same result. By construction, the probability distribution induced on this truncated words is exactly $\mathbb{R}_n$, concluding the proof.

The generalisation to cyclically reduced words follows from Lemma 3.13. □

We now use Proposition 4.26 to prove Proposition 4.22.

Proof of Proposition 4.22 Let $0 < \lambda < \frac{1}{2}$. Proposition 4.26, applied to $\ell_n = \lambda n$ shows that, if $\frac{1}{2} < d < 1$, then a random $\alpha_{[2]}^{-d\lambda n}$-tuple $\vec{h}$ of cyclically reduced words of length n, generically has two components h_i and h_j with the same prefix of length λn, which is sufficient to show that $\vec{h}$ does not satisfy Property $C'(\lambda)$. □

We now translate the result of Proposition 4.26 into a result on the group presented by a random $\alpha_{[2]}^{-dn}$-tuple, when $d > \frac{1}{2}$. We will use repeatedly Chernoff bounds [**20**, Th. 4.2 p.70], which state that, in a binomial distribution with parameters (k, p) — that is: X_k is the sum of k independent draws of 0 or 1 and p is the

probability of drawing 1 —,

$$\mathbb{P}\left(X_k \leq \frac{kp}{2}\right) \leq \exp\left(-\frac{kp}{8}\right).$$

In other words,

$$\mathbb{P}\left(X_k \geq \frac{kp}{2}\right) \geq 1 - \exp\left(-\frac{kp}{8}\right). \tag{8}$$

If $\vec{h}$ is a vector of cyclically reduced words, G is the group presented by $G = \langle A \mid \vec{h}\rangle$ and u, v are reduced words, we write that $u =_G v$ if u and v have the same projection in G (that is: if uv^{-1} lies in the normal closure of $\vec{h}$).

PROPOSITION 4.27. *Let $\mathcal{A}$ be an ergodic Markovian automaton with coincidence probability $\alpha_{[2]}$ and let $a, b \in \tilde{A}$ be labels of transitions in $\mathcal{A}$. Let $d > \frac{1}{2}$ and $N \geq \alpha_{[2]}^{-dn}$, and let G be a group presented by an N-tuple of cyclically reduced words chosen at random according to $\mathcal{A}$. Then $a =_G b$ exponentially generically.*

PROOF. Let $t > 0$ be such that all the coefficients of $\mathbb{M}^t$ are positive (such an integer exists since $\mathbb{M}$ is ergodic) and let $\tau > 0$ be the minimum coefficient of $\mathbb{M}^t$.

We proceed in two steps. First we consider transitions starting in the same state of the Markovian automaton and second we generalize the study to transitions beginning in different states of the automaton.

First step of the proof. We show that if $x = x_1 \cdots x_s$ and $y = y_1 \cdots y_s$ are reduced words of equal length $s \geq 1$ which label paths in $\mathcal{A}$ out of the same state q, then exponentially generically, we have $x_k =_G y_k$ for each $1 \leq k \leq s$.

Recall that, in our model of Markovian automata, drawing a word of length n amounts to drawing a state $r \in Q$ according to γ_0, and then drawing a word of length n according to $\gamma(r, -)$. Thus, when drawing a tuple $\vec{h} = (h_i)_i$, we also draw a tuple $\vec{q} = (q_i)_i$ of states such that, in particular, $\gamma_0(q_i) > 0$ and $\gamma(q_i, h_i) > 0$.

Let r be a state such that $\gamma_0(r) > 0$. Let $T_0 = \{h_i \in \vec{h} \text{ such that } q_i = r\}$ and $N_0 = |T_0|$. Observe that drawing randomly and independently N words of length n in our model and then keeping only those starting in state r to obtain T_0 is the same as first choosing N_0 according to a binomial law of parameters $(\gamma_0(r), N)$ and then drawing randomly and independently N_0 words beginning in state r. Moreover Chernoff bounds (Equation (8) above, applied with $p = \gamma_0(r)$ and $k = N$) show that $\mathbb{P}\left(N_0 \geq \frac{\gamma_0(r)N}{2}\right) \geq \mathfrak{p}_0$ with $\mathfrak{p}_0 = 1 - \exp\left(-\frac{\gamma_0(r)N}{8}\right)$.

For each $s \geq 1$, we say that a pair of indices (i, j) is an *s-collision in T_0* if h_i and h_j belong to T_0 and have the same prefix of length $n - t - s$. Let e be such that $0 < e < d - \frac{1}{2}$ and let $N' = \alpha_{[2]}^{-(d-e)n}$. Then a random N_0-tuple of cyclically reduced words starting in r is obtained by drawing $\frac{N_0}{N'}$ times a random N'-tuple starting in state r. Moreover choosing a random word in a Markovian automaton given that the associated path begins in state r is the same as taking for initial probability vector γ_0 the probability vector such that $\gamma_0(r) = 1$. Since the conclusion of Proposition 4.26 does not depend on the initial probability vector and $d - e > \frac{1}{2}$, Proposition 4.26 applied to $\ell_n = n - t - s$ shows that a random N'-tuple of cyclically reduced words that starts in r generically exhibits at least one s-collision in T_0.

We assume that n is large enough so that the probability of an s-collision in T_0 of a random N'-tuple is at least $\frac{1}{2}$. Then Chernoff bounds (Equation (8), applied

with $p = \frac{1}{2}$ and $k = N_0$) show that the set T_1 of s-collisions in T_0 of a random $\alpha_{[2]}^{-dn}$-tuple of cyclically reduced words of length n satisfies $|T_1| \geq \frac{1}{4}N_0$ with probability greater than or equal to $\mathfrak{p}_1 = 1 - \exp(-\frac{N_0}{16})$.

For each s-collision $(i,j) \in T_1$, we let $u(i,j)$ be the common length $n-t-s$ prefix of h_i and h_j. Then by a finiteness argument, there exists a state $q_1 \in Q$ and a set $T_2 \subset T_1$ such that, for every $(i,j) \in T_2$, $u(i,j)$ labels a path from r to q_1 in $\mathcal{A}$, and $|T_2| \geq \frac{|T_1|}{|Q|}$. Hence $|T_2| \geq \frac{N_0}{4|Q|}$ with probability greater than or equal to $\mathfrak{p}_1$.

Now let v be a reduced word of length t, labeling a path in $\mathcal{A}$ from q_1 to q: such a word exists since all the coefficients of $\mathbb{M}^t$ are positive, and we have $\gamma(q_1, v) \geq \tau$. For each $(i,j) \in T_2$, the probability that h_i starts with $u(i,j)v$ is $\gamma(q_1, v) \geq \tau$, and the probability that uv is a prefix of both h_i and h_j is at least τ^2. We can apply Chernoff bounds (8) again, with $p = \tau^2$ and $k = |T_2|$: then the subset $T_3 \subseteq T_2$ of pairs (i,j) such that $u(i,j)v$ is a prefix of both h_i and h_j, has cardinality $|T_3| \geq \frac{1}{2}|T_2|\tau^2$ with probability at least $\mathfrak{p}_2 = 1 - \exp(-\frac{\tau^2|T_2|}{8})$.

Finally, we note that $|u(i,j)v| = n - s$, so for each $(i,j) \in T_3$, we have $h_i = u(i,j)vx$ with probability $\gamma(q,x)$. Therefore the probability that $(h_i, h_j) = (u(i,j)vx, u(i,j)vy)$ is $\gamma(q,x)\gamma(q,y)$, which is positive by hypothesis. Applying Chernoff bounds one more time (with $k = |T_3|$ and $p = \gamma(q,x)\gamma(q,y)$) shows that $\vec{h}$ contains a pair of words of the form (wx, wy) with probability at least $\mathfrak{p}_3$ with $\mathfrak{p}_3 = \left(1 - \exp\left(-\frac{|T_3|\gamma(q,x)\gamma(q,y)}{8}\right)\right)$.

In conclusion, exponentially generically $N_0 \geq \gamma_0(r)\frac{N}{2}$ which implies that $\mathfrak{p}_1$ is exponentially close to 1. Hence $T_2 \geq \frac{\gamma_0(r)N}{8|Q|}$ exponentially generically, which implies that $\mathfrak{p}_2$ is exponentially close to 1. So $|T_3| \geq \frac{\gamma_0(r)N\tau^2}{16|Q|}$ exponentially generically, which implies that $\mathfrak{p}_3$ is exponentially close to 1. In particular, exponentially generically, $\vec{h}$ has a pair of the form (wx, wy), and hence we have $x =_G y$.

Applying this to the words x_1 and y_1, we find that $x_1 =_G y_1$. Next, considering the words x_1x_2 and y_1y_2, we find that $x_1x_2 =_G y_1y_2$, and hence $x_2 =_G y_2$. Iterating this reasoning, we finally show that $x_k =_G y_k$ for each $1 \leq k \leq s$.

Second step of the proof We now consider two transitions in $\mathcal{A}$, one labeled a from state q to state q' and another labeled b from state r to state r' $(a, b \in \tilde{A})$.

Let $q_0 \in Q$ be a state in $\mathcal{A}$ such that $\gamma_0(q_0) > 0$. Since $\mathcal{A}$ is irreducible, there exists a word w_1 which labels a loop at q_0 and visits every transition of $\mathcal{A}$. Moreover, since $\mathcal{A}$ is ergodic, there exists a word w_2 labeling another loop at q_0, such that $|w_1|$ and $|w_2|$ are relatively prime.

Since reading w_1 from q_0 visit all the transitions, let u_1 (resp. v_1) be a prefix of w_1 such that the last transition read after reading u_1 (resp. v_1) is the a-transition out of state q (resp. the b-transition out of state r). Then the Chinese remainder theorem shows that there exist words $x \in \{w_1, w_2\}^*u_1$ and $y \in \{w_1, w_2\}^*v_1$ of equal length.

Since a and b are the last letters of x and y, respectively, the first step of the proof shows that $a =_G b$, which concludes the proof of the proposition. □

We can now complete the **proof of Proposition 4.23.** By Proposition 4.27, exponentially generically, all the letters in E are equal in G. If $a, a^{-1} \in E$ for some letter a, then all these letters are equal to their own inverse in G, so the subgroup H of G generated by E is a quotient of $\mathbb{Z}/2\mathbb{Z}$. Since all the relators in the presentation

have length n, it follows that H is isomorphic to $\mathbb{Z}/2\mathbb{Z}$ if n is even, and is trivial if n is odd. The result follows once we observe that the letters in D do not occur in any relator. □

Acknowledgments. The authors are thankful to the anonymous referee for her/his remarkably thorough reading of the first version of this paper and for his/her insightful and constructive suggestions. These helped simplify the presentation of Sections 3.2 and 3.3, sharpen some results in Section 3.6 and fix a technical mistake in the proof of Proposition 4.23.

References

[1] G. N. Arzhantseva and A. Y. Ol'shanskiĭ. Generality of the class of groups in which subgroups with a lesser number of generators are free. *Mat. Zametki*, 59(4):489–496, 638, 1996.

[2] F. Bassino, C. Nicaud, and P. Weil. Generic properties of random subgroups of a free group for general distributions. In *23rd Intern. Meeting on Probabilistic, Combinatorial, and Asymptotic Methods for the Analysis of Algorithms (AofA'12)*, Discrete Math. Theor. Comput. Sci. Proc., AQ, pages 155–166. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2012.

[3] F. Bassino, A. Martino, C. Nicaud, E. Ventura, and P. Weil. Statistical properties of subgroups of free groups. *Random Structures Algorithms*, 42(3):349–373, 2013.

[4] F. Bassino, C. Nicaud, and P. Weil. Random generation of finitely generated subgroups of a free group. *Internat. J. Algebra Comput.*, 18(2):375–405, 2008.

[5] F. Bassino, C. Nicaud, and P. Weil. On the genericity of Whitehead minimality. *J. Group Theory*, to appear, 2015.

[6] M. R. Bridson and D. T. Wise. Malnormality is undecidable in hyperbolic groups. *Israel J. Math.*, 124:313–316, 2001.

[7] C. Champetier. Propriétés statistiques des groupes de présentation finie. *Journal of Advances in Mathematics*, 116(2):197–262, 1995.

[8] F. R. Gantmacher. *The theory of matrices*. Chelsea, 1959.

[9] R. Gitik, M. Mitra, E. Rips, and M. Sageev. Widths of subgroups. *Trans. Amer. Math. Soc.*, 350(1):321–329, 1998.

[10] M. Gromov. Hyperbolic groups. In *Essays in group theory*, volume 8 of *Math. Sci. Res. Inst. Publ.*, pages 75–263. Springer, New York, 1987.

[11] M. Gromov. Asymptotic invariants of infinite groups. In *Geometric group theory, Vol. 2 (Sussex, 1991)*, volume 182 of *London Math. Soc. Lecture Note Ser.*, pages 1–295. Cambridge Univ. Press, Cambridge, 1993.

[12] T. Jitsukawa. Malnormal subgroups of free groups. In *Computational and statistical group theory (Las Vegas, NV/Hoboken, NJ, 2001)*, volume 298 of *Contemp. Math.*, pages 83–95. Amer. Math. Soc., Providence, RI, 2002.

[13] I. Kapovich. Musings on generic-case complexity. `arXiv:1505.03218`, 2015.

[14] I. Kapovich, A. Miasnikov, P. Schupp, and V. Shpilrain. Generic-case complexity, decision problems in group theory, and random walks. *J. Algebra*, 264(2):665–694, 2003.

[15] I. Kapovich and A. Myasnikov. Stallings foldings and subgroups of free groups. *J. Algebra*, 248(2):608–668, 2002.

[16] S. Karlin and F. Ost. Counts of long aligned word matches among random letter sequences. *Adv. in Appl. Probab.*, 19(2):293–351, 1987.

[17] O. Kharlampovich and A. Myasnikov. Hyperbolic groups and free constructions. *Trans. Amer. Math. Soc.*, 350(2):571–613, 1998.

[18] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009.

[19] R. C. Lyndon and P. E. Schupp. *Combinatorial group theory*. Springer-Verlag, Berlin, 1977. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 89.

[20] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

[21] A. Miasnikov, E. Ventura, and P. Weil. Algebraic extensions in free groups. In *Geometric group theory*, Trends Math., pages 225–253. Birkhäuser, Basel, 2007.

[22] J. Nielsen. Die Isomorphismen der allgemeinen, unendlichen Gruppe mit zwei Erzeugenden. *Mathematische Annalen*, 78, 1918.

[23] Y. Ollivier. Sharp phase transition theorems for hyperbolicity of random groups. *Geom. Funct. Anal.*, 14(3):595–679, 2004.
[24] Y. Ollivier. *A January 2005 invitation to random groups*, volume 10 of *Ensaios Matemáticos [Mathematical Surveys]*. Sociedade Brasileira de Matemática, 2005.
[25] A. Y. Ol'shanskiĭ. Almost every group is hyperbolic. *Internat. J. Algebra Comput.*, 2(1):1–17, 1992.
[26] M. O. Rabin. Probabilistic automata. *Information and Computation*, 6(3):230–245, 1963.
[27] A. Roig, E. Ventura, and P. Weil. On the complexity of the Whitehead minimization problem. *Internat. J. Algebra Comput.*, 17(8):1611–1634, 2007.
[28] R. Segala and N. Lynch. Probabilistic simulations for probabilistic processes. *Nordic Journal of Computing*, 2(2):250–273, 1995.
[29] J.-P. Serre. *Trees*. Springer-Verlag, Berlin, 1980. Translated from the French by John Stillwell.
[30] P. V. Silva and P. Weil. On an algorithm to decide whether a free group is a free factor of another. *Theor. Inform. Appl.*, 42(2):395–414, 2008.
[31] J. R. Stallings. Topology of finite graphs. *Invent. Math.*, 71(3):551–565, 1983.
[32] W. Szpankowski. On the height of digital trees and related problems. *Algorithmica*, 6(2):256–277, 1991.
[33] N. W. M. Touikan. A fast algorithm for Stallings' folding process. *Internat. J. Algebra Comput.*, 16(6):1031–1045, 2006.
[34] B. Vallée, J. Clément, J. A. Fill and Ph. Flajolet. The number of symbol comparisons in QuickSort and QuickSelect. *Automata, Languages and Programming*, Springer, pages 750–763. Berlin Heidelberg, 2009.
[35] P. Weil. Computing closures of finitely generated subgroups of the free group. In *Algorithmic problems in groups and semigroups (Lincoln, NE, 1998)*, Trends Math., pages 289–307. Birkhäuser Boston, Boston, MA, 2000.

Université Paris 13, Sorbonne Paris Cité, LIPN, CNRS UMR 7030, F-93430 Villetaneuse, France
E-mail address: bassino@lipn.univ-paris13.fr

Université Paris-Est, LIGM (UMR 8049), UPEMLV, F-77454 Marne-la-Vallée, France
E-mail address: nicaud@univ-mlv.fr

Univ. Bordeaux, LaBRI, CNRS UMR 5800, F-33400 Talence, France
E-mail address: pascal.weil@labri.fr